# The interplay between genes and dietary factors in the aetiology of Type 2 Diabetes Mellitus

**Sherly (Xueyi) Li**

MRC Epidemiology Unit and Lucy Cavendish College

University of Cambridge

This dissertation is submitted for the degree of

*Doctor of Philosophy*

August 2017

# *Acknowledgements*

Whilst undertaking this PhD, I have been privileged to work with talented yet humble and nurturing people, who have shaped me both professionally and personally.

I was new to the disciplines of genetics and epidemiology when starting my PhD so the MRC Epidemiology Unit provided a stimulating and supportive learning environment. Scientists like Fumiaki Imamura, Felix Day and John Perry not only challenged my thinking, but they invested time in teaching and mentoring students, like me, which I am grateful for.

The opportunity to work across two groups within our unit (Nutritional Epidemiology and Diabetes Aetiology) allowed me to learn the best from both worlds. It broadened my insights from different perspectives but also challenged me to develop simpler ways of explaining complex concepts and methods between disciplines, appreciate differences in understanding, resolving problems and advocate for greater mutual understanding and sharing of knowledge between disciplines. I am grateful to the following people for providing their unique perspectives, expertise and for stimulating thoughtful discussions: Zheng Ye, Laura O'Connor, Simon Wheeler, Jusheng Zheng, Anouar Fanidi, Matthew Pearce, Silvia Pastorino, Birdem Amoutzopoulos, Marinka Steur, Jing Hua Zhao, Luca Lotta, Alexia Cardona and Isobel Stewart. I'd like to thank Nicola Kerrison for her helpfulness with EPIC-InterAct and Jian'an Luan and Stephen Sharp for their statistical guidance. Thank you also to Meriel Smith for her resourcefulness and support. A significant source of scientific inspiration also came from small

corridor chats and in having role models among my fellow PhD students. So, thank you Laura Wittemans, Chen Li, Clara Podmore, Eirini Trichia, Tom White, Dan Wright, Emma Clifton, Sundas Javad, Tammy Tong, Maxine Lamb and Tarra Penney.

There are three women who have, throughout my PhD, been invaluable mentors and friends. Barbora Silarova, Sara Willems and Adina Feldman shared experiences from their PhDs to help me through mine, inspired my thinking, fostered emotional strength, and provided constructive support in my ideas and initiatives.

I cannot thank enough my family, my mum and dad, and close friends for their love and persistent faith in me. My resilience and optimism come from having people like these in my life.

Lastly, I have been extremely lucky to have worked with people who have professionally inspired me and championed me towards my aspirations. Helen Truby and Claire Palermo (Monash University, Australia) and Kevin Whelan (King's College London) first introduced me to the world of research and offered me the opportunities that led me to undertaking this PhD. This I will be forever grateful for. During my PhD, my supervisors, Nita Forouhi, Nick Wareham and Robert Scott, instilled a sense of confidence in me for recognising the importance of honest and rigorously conducted research, regardless of the findings. I'd like to thank them for their generous guidance and support over the last 3 years, especially in teaching me the principles of good research practice. A special thanks to Nita, my primary supervisor, for her scientific insights and mentorship. She helped to develop my critical thinking, was patient with me, always made the time to provide thorough and constructive feedback to my work and offered me invaluable opportunities throughout my PhD. I know I will continue to be inspired by these incredible people who I admirable not only for their professional success but dedication and kindness in how they conduct themselves.

Thank you all for shaping the person I am today.

# *Declarations*

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where detailed in the Contributions section on p.g.10, projects table on p.g.201, or specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently
submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of a similar institution.

This dissertation does not exceed 60,000 words.

Sherly (Xueyi) Li

# *Summary of Thesis*

To help mitigate the escalating prevalence of Type 2 Diabetes (T2D) and alleviate society of its associated morbidity and economic burden on health care, it is crucial to understand its aetiology. Both genetic and the environmental risk factors are known to be involved. Healthy diets have been proven to reduce the risk of T2D in primary prevention trials, however, which components and exact mechanisms are involved is not fully understood, in particular, the role of macronutrient intake. Body weight, glycaemic markers and T2D are all to some extent genetically regulated. There may also be genetic influences on how people digest, absorb or metabolise macronutrients. This poses the possibility that the interplay between genes and our diet may help us unravel T2D's aetiology.

The aim of this PhD was to investigate gene-diet interactions on the risk of incident T2D, focusing primarily on macronutrient intake as the dietary factor. First, I systematically evaluated the current evidence before taking a step-wise approach (hypothesis driven to hypothesis-free) to interrogate gene-macronutrient interactions. This identified 13 publications, with 8 unique interactions reported between macronutrients (carbohydrate, fat, saturated fat, dietary fibre, and glycaemic load derived from self-report of dietary intake and circulating n-3 polyunsaturated fatty acids) and genetic variants in or near *TCF7L2*, *GIPR*, *CAV2* and *PEPD* ($p<0.05$) on T2D. All studies were observational with moderate to serious risk of bias and limitations that included lack of adequate adjustment for confounders, lack of reported replication and insufficient correction for multiple testing.

Second, these reported interactions did not replicate in a large European multi-centre prospective T2D case-cohort study called EPIC-InterAct. We concluded that the heterogeneity between our results and those published could be explained by methodological differences in dietary measurement, population under study, study design and analysis but also by the possibility of spurious interactions.

Third, given the paucity of gene-macronutrient interaction research using genetic risk scores (GRS), we examined the interaction between three GRS (for BMI (97 SNPs), insulin resistance (53 SNPs) and T2D (48 SNPs)) and macronutrient intake (quantity and quality indicators) in EPIC-InterAct. We did not identify any statistically significant interactions that passed multiple testing corrections ($p \geq 0.20$, with a p value threshold for rejecting the null hypothesis of 0.0015 (based on 0.05/33 tests)). We also examined 15 foods and beverages identified as being associated with T2D, and no significant interactions were detected. Lastly, we applied a hypothesis-free method to examine gene-macronutrient interactions and T2D risk by using a genome-environment-wide-interaction-study. Preliminary findings showed no significant interactions for total carbohydrate, protein, saturated fat, polyunsaturated fat and cereal fibre intake on T2D.

In conclusion, the consistently null findings in this thesis using a range of statistical approaches to examine interactions between genetic variants and macronutrient intake on the risk of developing T2D have two key implications. One, based on the specific interactions examined, this research does not confirm evidence for gene-diet interactions in the aetiology of T2D and two, this research suggests that the association between macronutrient intake and the risk of developing T2D does not differ by genotype.

# *Publications related to this thesis*

**Li SX**, Ye Z, Whelan K, Truby H. The effect of communicating the genetic risk of cardiometabolic disorders on motivation and actual engagement in preventative lifestyle modification and clinical outcome: a systematic review and meta-analysis of randomised controlled trials. *Br J Nutr* 2016;116(5):924–34

**Li SX**, Imamura F, Ye Z, Schulze MB, Zheng J, Ardanaz E, et al., Interaction between genes and macronutrient intake on the risk of developing type 2 diabetes: systematic review and findings from EPIC- InterAct. *Am J Clin Nutr* 2017;106(1):263-75

**Li SX**, Imamura F, Ye Z, Schulze MB, Zheng J, Ardanaz E, et al., Interplay between genetic predisposition and macronutrient intake on type 2 diabetes incidence: analysis within EPIC-InterAct across eight European countries. *Diabetologia* 2018. doi: 10.1007/s00125-018-4586-2. [Epub ahead of print]

**Li SX**, Imamura F, Ye Z, Schulze MB, Zheng J, Ardanaz E, et al., Interplay between genetic predisposition, food and beverage intake and Type 2 Diabetes incidence. Analysis across 8 European countries: EPIC-InterAct Study. *Manuscript under review*

**Li SX**, Imamura F, Ye Z, Schulze MB, Zheng J, Ardanaz E, et al., Genome-wide-interaction study of macronutrient intake and incident Type 2 Diabetes: EPIC-InterAct Study. *Project on-going*

*I have only listed publications arising from projects that I have led on. Others may arise from projects I have contributed to (please see Table on p.g. 201).*

# *Presentations related to this thesis*

## Oral presentations

Is personalised nutrition overhyped? Lucy Cavendish College graduate research day (7-5-2015)

Could personalised nutrition, based on genetics, be possible for Type 2 Diabetes Mellitus? Rank Prize Symposium, Lake District England (19-10-2015)

Substitution of macronutrients, gene-macronutrient interaction and development of Type 2 Diabetes. Danish Nutrition Society Meeting, Denmark (28-2-2017)

## Poster presentations

Are we ready for personalised nutrition? Interaction between macronutrient intake and genetic factors on the risk of type 2 diabetes: a systematic review. Public Health @ Cambridge, Cambridge England (8-6-2015)

Interaction between genes and macronutrient intake on the risk of developing type 2 diabetes: systematic review and findings from EPIC-InterAct. Institute of Metabolic Science Away Day, Cambridge England (13-11-2015)

Interaction between genes and macronutrient intake on the risk of developing type 2 diabetes: systematic review and findings from EPIC-InterAct. Precision Medicine Symposium, Sweden (30-8-2016)

Will knowing your genetics change your behaviour? International Congress of Dietetics, Spain (7-9-2016)

## *Other publications*

**Li SX**, Day F. Mendelian Randomisation in nutritional epidemiology: what dietitians need to know? *PEN Methods collection and eNews* published on 2017 April 17th

Abrahams M, Adamski M, **Li SX**. Knowledge pathway on nutritional genomics (*PEN*: Practice-based Evidence in Nutrition. These knowledge pathways serve as quick references to evidence-based information to help guide clinical decision making for dietitians) *Guideline under review*

Whelan K, **Li SX**. Genetics and Nutritional Genomics. In: Gandy J, editor. *Manual of Dietetic Practice, 6th Edition.* (A core undergraduate textbook for dietitians) *Under review*

# *Contributions*

This thesis and related publications entail the input of many scientists involved in EPIC-InterAct study. The contributions that they have made to the projects that I have led on, detailed in Chapters 4 to 7, are described below.

My supervisors, Nita Forouhi, Robert Scott and Nick Wareham provided advice and direction for these projects, contributed to interpretation of results and critically reviewed and edited the manuscripts. Nita Forouhi and Nick Wareham also provided advice and feedback on the writing of this thesis.

Fumiaki Imamura, Zheng Ye, Jusheng Zheng and Matthias Schulze were the working group members for the projects that I led on, contributing to the development of the analysis plan and critical review and editing of the manuscripts. The wider InterAct consortium also provided critical review and editing of the manuscript for Chapter 4 and 5.

The EPIC-InterAct project was coordinated by Nick Wareham, Claudia Langenberg, Nita Forouhi and Stephen Sharp and in addition to this; they contributed to the review of Chapters 4 and 5.

My supervisors, Fumiaki Imamura, Sara Willems and Felix Day provided guidance on the plans for the genome-environment-wide-interaction-study (GEWIS, Chapter 7). Stephen Sharp and Jian'an Luan provided statistical advice. Jian'an Luan also performed the statistical analyses for the GEWIS.

Nicola Kerrison provided overall data management for the EPIC-InterAct study.

I led on the projects described in Chapters 4 to 7, in terms of undertaking the systematic review (pertaining to Chapter 4), performing the statistical analyses, interpretation of the findings and wrote the first draft of the manuscripts. I also undertook post-analysis Quality Control and meta-analyses of the genotype chip specific GEWIS results (pertaining to Chapter 7).

In the discussion chapter, I have included work that I have contributed to as part of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Nutrition Group.

# Table of Contents

# *Index of Figures*

# *Index of Tables*

# *Abbreviations*

| | |
|---|---|
| ACCE | analytic validity, clinical validity, clinical utility and associated ethical, legal and social implications |
| BMI | Body Mass Index |
| CHARGE | Cohorts for Heart and Aging Research in Genomic Epidemiology |
| CI | Confidence Interval |
| CVD | Cardiovascular Disease |
| DAG | Directed Acyclic Graph |
| DTC | Direct-To-Consumer |
| EPIC | European Prospective Investigation into Cancer and Nutrition |
| FFQ | Food Frequency Questionnaire |
| GEWIS | Genome-environment-wide-interaction-study |
| GRS | Genetic Risk Score |
| GWAS | Genome-wide-association-study |
| HOMA-IR | Homeostatic Model Assessment- Insulin Resistance |
| HR | Hazard Ratio |
| HWE | Hardy-Weinberg Equilibrium |
| IR | Insulin Resistance |
| LD | Linkage Disequilibrium |
| MAF | Minor Allele Frequency |
| MICE | Multiple Imputation by Chained Equations |

| | |
|---|---|
| MR | Mendelian Randomisation |
| MUFA | Monounsaturated fatty acid |
| OR | Odds Ratio |
| PhD | Doctorate of Philosophy |
| PUFA | Polyunsaturated fatty acid |
| QQ | Quintile-Quintile |
| RCT | Randomised Control Trial |
| RHR | Relative Hazard Ratio |
| RR | Relative Risk |
| SE | Standard Error |
| SFA | Saturated fatty acid |
| SMD | Standardised Mean Difference |
| SNP | Single Nucleotide Polymorphism |
| SSB | Sugar Sweetened Beverage |
| T2D | Type 2 Diabetes |
| TEI | Total Energy Intake |
| UK | United Kingdom |

# *Chapter 1 Introduction*

In 1932, Lancelot Hogben a zoologist and medical statistician claimed that variability

> '*arises from the combination of a particular hereditary constitution with a particular kind of environment*.'[1]

The variability in disease traits and the contribution of nature vs nurture to this variability has fascinated scientists for centuries. Certainly, for Type 2 Diabetes Mellitus (T2D), a major global public health challenge of the 21st century, variability in development, manifestation and progression have underpinned the many discussions about its definition and classification.[2] This is important in correct and timely diagnosis and treatment but also understanding what drives variability in disease development offers crucial insights to better target prevention efforts. Therefore, to better understand this variability and T2D aetiology, it led me to study the interplay between our genes and dietary intake (i.e. interplay between nature and nurture).

Apart from helping to unravel the aetiology of T2D, it is believed that knowledge about gene-diet interactions may translate into identifying individuals most likely to benefit from preventative dietary interventions.[3] This is of great public health interest, particularly in the setting of scarce healthcare funding. However currently, it is unclear what the evidence is for gene-diet interactions in the development of T2D. That is, are there any convincing interactions that may support this ambition? Studies suggesting gene-diet interactions for T2D certainly exist, including between dietary fibre and genetic predisposition for T2D via *TCF7L2*[4] (strongest genetic locus for T2D),[5] however, findings have been sporadic and inconsistent. Secondly, the literature on gene-environment

interactions harbours many methodological limitations that devalue the quality of studies used to evaluate the evidence. This includes problems inherent to the observational nature of studies, such as chance, confounding and bias, as well as concerns regarding the lack of statistical power because of the relatively small magnitude of interaction effects and publication bias. Hence, would studies designed to overcome such limitations provide a clearer answer?

The overall aim of this PhD thesis was therefore to address these two questions about gene-diet interactions in the context of preventing T2D, in order to understand if these are important in the aetiology of T2D as well as the value for dietary public health interventions. This was achieved using observational epidemiological methods applied to large-scale population-based and T2D case-cohort studies.

This thesis begins by introducing the background that motivated this piece of research in Chapter 2, ending with the aims and specific objectives. Next, the concepts and application of statistical interactions used within this PhD will be explained before a description of the main case-cohort study that was used within this PhD: EPIC-InterAct (Chapter 3). In Chapter 4, the literature on gene-macronutrient interactions and T2D was reviewed and to assess the validity of reported interactions an attempt was made to replicate these in EPIC-InterAct. Gaps in research were identified, one of which was a lack of non-candidate gene approaches to examine interactions. Therefore, in Chapter 5 and 6, genetic risk scores were used to investigate interactions with macronutrients and with foods and beverages in relation to the development of T2D. A genome-environment-wide-interaction-study to examine gene-macronutrient interactions was also conducted and will be discussed in Chapter 7.

Lastly, in Chapter 8, a discussion of the strengths and limitations will be made in the context of the current literature, including a discussion about the impact of genetic risk information on behaviour change which together with the findings on gene-diet interactions helps to evaluate the public health value of genetically stratified nutritional interventions. Finally, a discussion about what the findings mean for future research and public health nutrition draws the thesis to a close.

# *Chapter 2 Background and literature review*

This chapter highlights the urgent need to tackle the rising prevalence of Type 2 Diabetes Mellitus (T2D). I also explain how prevention strategies may be optimised by improving our understanding of whether the risk of diet-induced T2D differs genetically between people. I will examine the evidence which motivated the research within this PhD.

## *2.1 Definition, classification and diagnosis of Type 2 Diabetes Mellitus*

According to the World Health Organization (WHO) the term "diabetes mellitus" describes, 'a metabolic disorder of multiple aetiology characterised by chronic hyperglycaemia with disturbances of carbohydrate, fat and protein metabolism resulting from defects in insulin secretion, insulin action, or both.'[6] T2D is often distinguished from other forms of diabetes by β-cell deterioration or dysfunction and results in an imbalance between insulin secretion and insulin sensitivity. Individuals with T2D may not require external insulin to survive, as depicted in the figure below from the American Diabetes Association (ADA) (Figure 2-1).

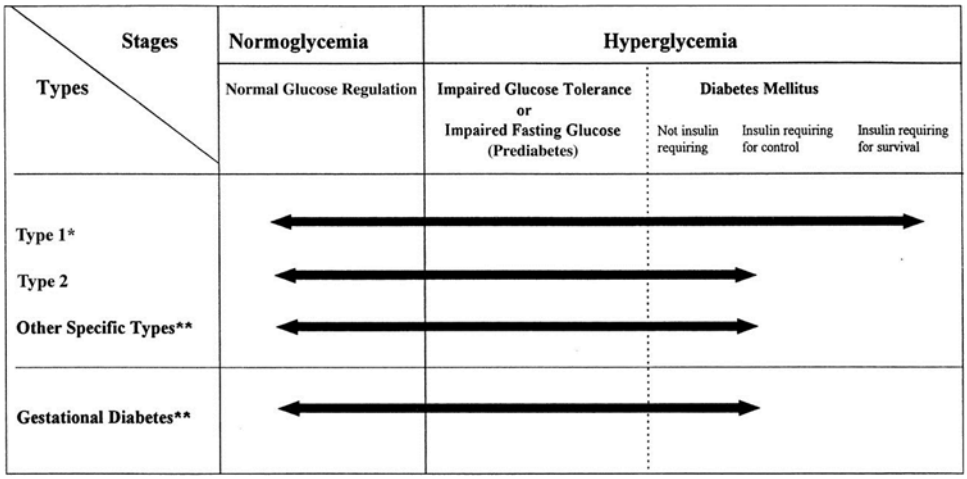



**Figure 2-1: Classification of diabetes (American Diabetes Association)**[7]

Diagnosing T2D involves either an examination of plasma glucose or haemoglobin A1c (HbA1c). With the former, this may be via fasting glucose or an oral glucose tolerance test followed by a plasma glucose test after 2 hours. The diagnostic criteria recommended by the WHO and the International Diabetes Federation (IDF) are provided in Table 2-1.[8,9] Concordance between the tests is imperfect and there are benefits and limitations of each. For example, HbA1c tests are the most stable to day-to-day perturbations and convenient but HbA1c is affected by factors that influence haemoglobin glycation independent of glycaemia, such as age, race and anaemia.[9] It appears that 2 hour plasma glucose during an oral glucose tolerance test diagnoses more people with diabetes. For diagnosing clinical diabetes, the ADA recommends that a second confirmatory test using a new blood sample is required when the patient is asymptomatic and test results exceed normal thresholds.[10] Classic symptoms of T2D include frequent urination, excessive thirst, fatigue, blurred vision and/or possible unexplained weight loss.

**Table 2-1: T2D diagnostic criteria**

| Test | Threshold above which diabetes is diagnosed |
|---|---|
| Fasting plasma glucose | ≥7mmol/L |
| 2 hour plasma glucose during an oral glucose tolerance test | ≥11.1mmol/L |
| HbA1c | ≥48mmol/mol (or ≥6.5%) |
| Patient with classical symptoms of hyperglycaemia, a random plasma glucose | ≥11.1mmol/L |

Even within T2D, based on current diagnostic criteria, presentation and the rate of progression of hyperglycaemia can vary between individuals. This heterogeneity may result in misdiagnosis or underdiagnosis in the clinical setting and/or misclassification for research purposes. Tuomi and colleagues described how the true spectrum and distinct forms of diabetes are still to be clarified and how factors such as the age of onset and genetics may contribute to this heterogeneity as well as may explain why there are patients presenting with features overlapping multiple types of diabetes.[2] Although this discussion is beyond the scope of this thesis, it suggests that there is much-needed research in dissecting the genetic contribution to this variation as it may help to elucidate a better understanding of the aetiology of T2D to improve the precision and timeliness of diagnosis and treatment.

## *2.2 Epidemiology of diabetes*

Worldwide prevalence of diabetes has more than doubled for men (from 4.3 to 9.0%) and increased by 60% for women (from 5.0 to 7.9%) from 1980 until 2014 and was estimated to affect more than 420 million people in 2014.[11] The fastest rise has been observed in low- and middle-income countries. In 2015, the average prevalence of diabetes in Europe was 9.1%, where 87-91% of these were cases of T2D.[12] Figure 2-2 illustrates the prevalence across Europe. According to the IDF, among European countries in 2015, 23.5 million people with diabetes (39.3%) were unaware of their condition, with T2D making up almost all those undiagnosed.[12] Untreated or poorly managed diabetes can contribute to significant morbidity, reduced quality of life and premature death, including microvascular complications such as retinopathy which increases the risk of blindness, nephropathy, neuropathy and macrovascular complications such as cardiovascular disease, which is the world's leading cause of mortality.[13] Women with gestational diabetes have higher risk of pregnancy complications such as pre-eclampsia and eclampsia but more devastating are the transgenerational impacts on the child. Studies have shown that children of women with gestational diabetes are at greater risk of developing T2D later in life because of shared genetics and intrauterine conditions that may cause 'fetal programming' due to

excessive placental transfer of glucose.[14] Every year, diabetes and its complications result in more than 2 million deaths, globally.[11] On top of this debilitating impact on human function and wellbeing, diabetes also poses immense economic burden. A recent systematic review estimated that the global economic cost of diabetes in 2015 was US$13.1 trillion, reflecting both direct (e.g. hospitalisation and medication) and indirect costs (e.g. loss of productivity due to morbidity and premature mortality).[15]

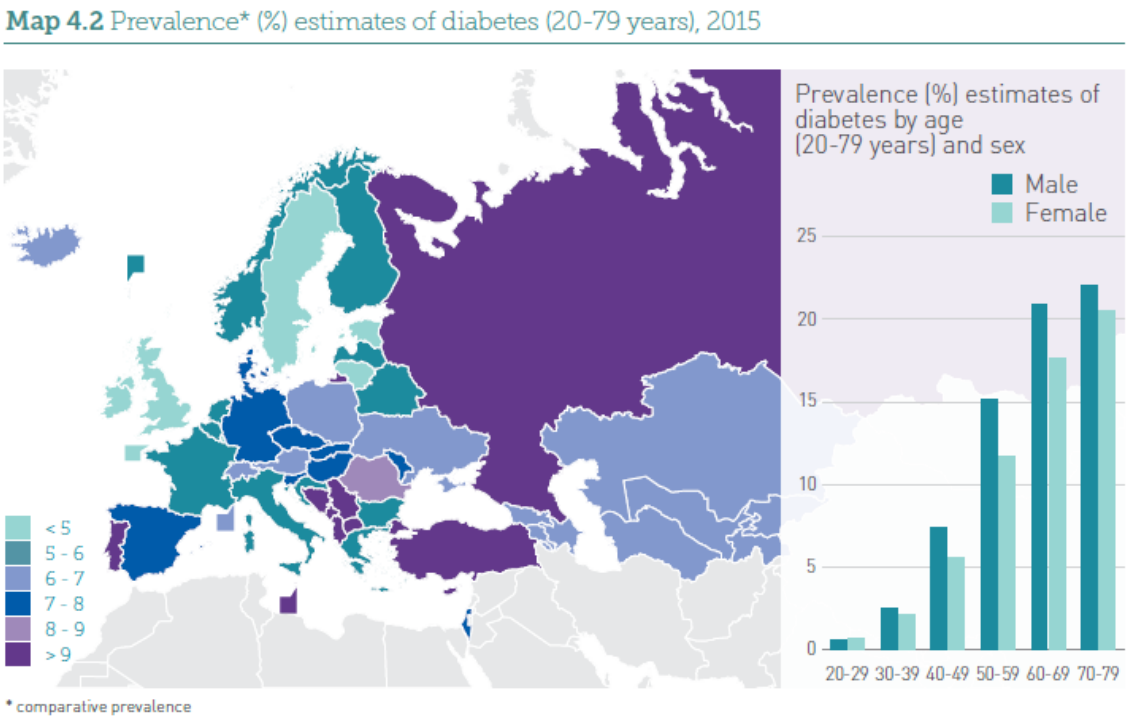



**Figure 2-2: Variation in prevalence of diabetes across Europe, IDF 2015**[12]

The sheer scale and disabling impacts of diabetes therefore identifies it as one of the four priority noncommunicable diseases within the 2011 Political Declaration on the Prevention and Control of Noncommunicable Diseases requiring urgent action by world leaders. It is also one of the key targets within the WHO Global Action Plan 2013-2020, which is a set of strategic actions towards the prevention and control of noncommunicable diseases.[16]

### *2.3 Pathophysiology and aetiology of type 2 diabetes*

T2D is multifactorial and polygenic and arises after the failure of insulin secretory function to maintain normoglycemia (β cell dysfunction) in the face of insulin resistance, often secondary to obesity.[17] Progressive β cell dysfunction characterises the glucose impairment classic of T2D.[17] This affects metabolism within three major tissues being the liver, muscle and adipose tissues (Figure 2-3).[18] Given that the brain and red blood cells' primary source of energy is glucose, the body tries to maintain sufficient levels of this. A lack of intracellular glucose signalling leads firstly to increased glucose production either by hepatic glycogenolysis or gluconeogenesis coupled with decreased glucose uptake by insulin-dependent GLUT4 of adipose tissues and muscle. This may lead to hyperglycaemia and subsequent microvascular conditions. Second, a reduction in carbohydrate utilisation may shift energy production from carbohydrate to another source such as fatty acids (increase lipolysis and decrease triglyceride synthesis) which may increase plasma lipids and macrovascular disease. Thirdly, muscle protein may be used as a source for gluconeogenesis and may lead to muscle wasting. Subsequently, hyperglycaemia and raised lipids may worsen insulin resistance.

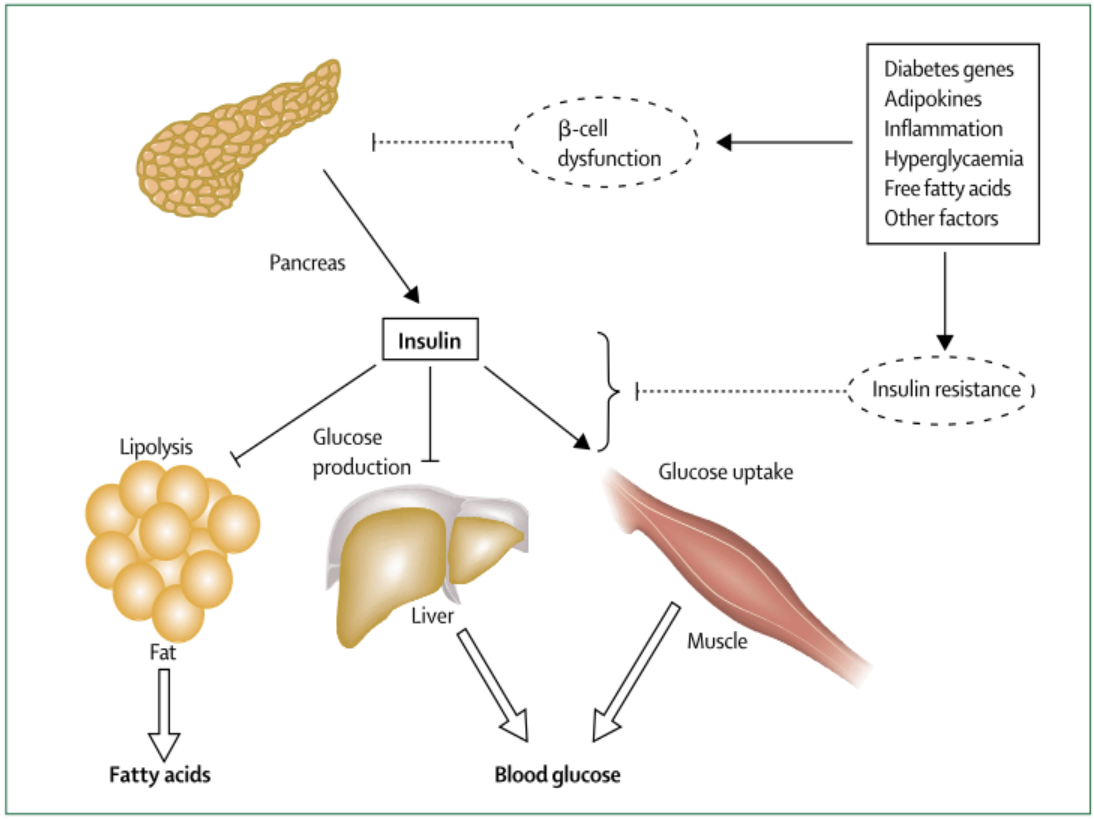



**Figure 2-3 Pathophysiology of hyperglycaemia and increased circulating fatty acids in type 2 diabetes**[18]

Fortunately, T2D can be prevented but in order to develop effective prevention strategies, we need to understand its aetiology. The risk factors for T2D and associated poor glucose homeostasis can be classified broadly as those that pertain to i) the individual's bio-physiology, ii) behaviours and iii) external influences of the environment or social context (Figure 2-4). There is overlap and interplay between these risk factors.

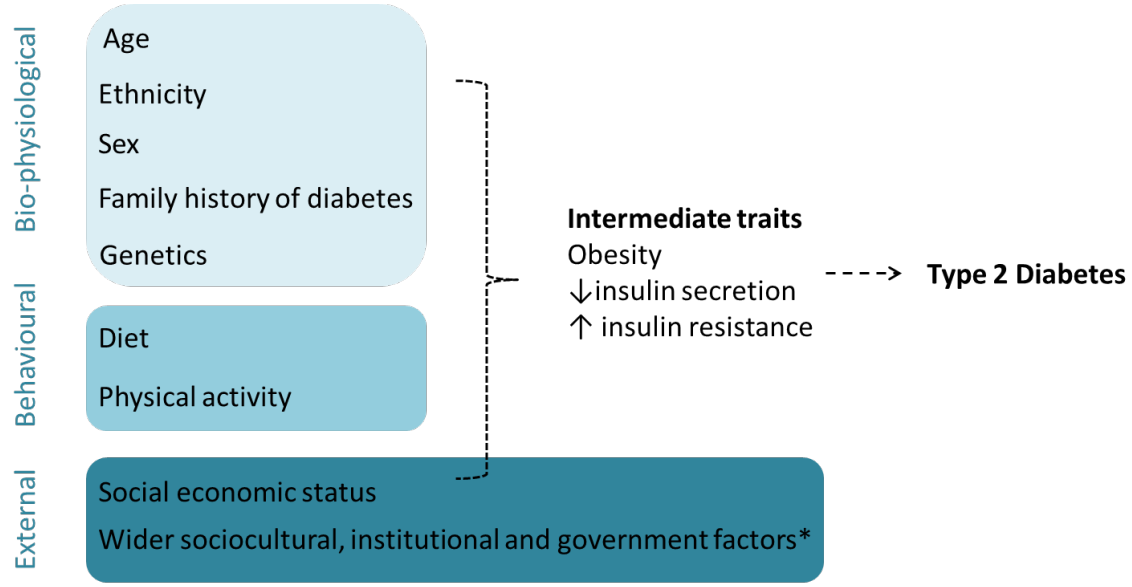



**Figure 2-4: Causal diagram of risk factors in the development of T2D (not an exhaustive list)**
* Represents factors illustrated in the Figure: Barriers and opportunities for healthy eating, available in Mozaffarian et al., 2016[19]

Within the domain of bio-physiological and behavioural factors, some are modifiable and some non-modifiable. Non-modifiable factors include ethnicity, age, sex, family history of diabetes and genetic predisposition, which will be discussed in section 2.4. Noticeably, a striking difference in diabetes prevalence exists between those under 40 and those over 50 years old, likely because T2D is heavily influenced by the accumulation of lifestyle factors over time. Such differences were also evident in the prevalence between men (9%) and women (7.9%) (Figure 2-2).[11,12]

Being overweight or obese drastically increases the risk of developing T2D. A meta-analysis of 11 cohort studies revealed significantly higher risk of T2D among those who were obese and metabolically healthy (defined by normal insulin sensitivity, normoglycaemia, low inflammation and higher cardiorespiratory fitness) (RR: 4.03, 95%CI: 2.66, 6.09) as well as those who were obese and metabolically unhealthy (RR: 8.93, 95%CI: 6.86, 11.62) compared to healthy normal-weight adults.[20] These individual factors (e.g. adiposity, blood glucose homeostasis, blood lipids and blood pressure) are responsive to lifestyle change (e.g. smoking, physical activity, sedentary time and diet) and thus are crucial factors to target for T2D prevention.

Wider societal influences including inequalities in access (financial, physical distance and comprehension), differing levels of acceptance and adherence to healthy lifestyles also affects risk of T2D. Those from socioeconomically disadvantaged backgrounds are particularly vulnerable to a poor diet and subsequent risk of T2D.[21] This was evident from a large UK-based cohort study, which found that individuals who consumed a greater diversity of foods (from five major food groups) had 30% lower risk of T2D than those consuming a diet with three or fewer food groups.[22] However, the group consuming a greater dietary diversity were often those who could afford to pay more.[22] These external factors, therefore, may either support healthy behaviours or create what some have coined an 'obesogenic environment,' that is not conducive to preventing T2D. These factors are the individual's social-cultural-geopolitical context and include the workplace, family environment, food production and industry interests, the country's nutrition policies, health care system, culture, and so on. This framework is in part underpinned by the Social-Ecological Model of health promotion, which recognises that human behaviours and well-being are a product of both their social, institutional and cultural context and the interaction with personal characteristics of the individual.[23]

It is through understanding the mechanisms of how our behaviours and environment shape our risk of T2D, in the context of our biological predispositions, that we can define more effective prevention strategies. Indeed, the importance of focusing on both individual and population level strategies is recognised in recommendations for prevention of T2D by the National Institute for Health and Care Excellence.[24] Among these many facets of risk factors involved in T2D development, my PhD focuses on the interplay between two of these, namely diet and genes (Figure 2-4). Next, these will be discussed separately before explaining the connection between the two.

### *2.4 Genetic epidemiology of T2D*

The heterogeneous development and clinical manifestation of T2D may in part be explained by biological differences between individuals, including their genes. Studies indicate that T2D is strongly heritable, with 61-78% of T2D variance

explained by genetics (estimated in >30,000 same-sex twin pairs aged ≥45 years)[25,26] with the genetic component for T2D often stronger than for type 1 diabetes.[12] Advancements in genetic epidemiology including recent meta-analyses of genome-wide association studies (GWAS) in >100,000 individuals have led to the discovery of many common genetic variants robustly associated with T2D (>70 single nucleotide polymorphisms- SNPs),[5,27] and related traits including 97 SNPs for body mass index,[28] 36 for fasting glucose, 19 for fasting insulin[29] and 53 for insulin resistance.[30] There is expected overlap in the genetic architecture of these traits, likely reflecting intermediate processes along the causal pathway for T2D.[31] The most commonly studied class of genetic variations are SNPs, which result in single nucleotide changes in the DNA sequence. For example, rs7903146 is a SNP in the *TCF7L2* gene, which encodes a high mobility group box-containing transcription factor involved in the Wnt signalling pathway, and is to date the most strongly associated SNP with T2D across multiple ethnicities (estimated odds ratio- OR per risk allele=1.45).[32,33] The majority of SNPs have smaller effects, with OR for T2D less than 1.20 per risk allele.[5]

Taken together, these genes only account for approximately 6% of the variance in T2D susceptibility,[5] which beg the question about missing heritability in further understanding the aetiology of T2D. We now know that low-frequency variants (0.1%<minor allele frequency<5%) play a smaller role in T2D compared to their common counterparts (2.9% vs 6.3% of the variance for T2D by rare vs common variants, respectively).[34] Therefore gene-environment interactions, including with our diet, are hypothesised to explain some of this missing heritability and has been of immense research interest.[35] Next, the evidence for diet will be appraised before the literature relating to gene-diet interactions will be summarised.

### *2.5 Diet and the prevention of Type 2 Diabetes*

Lifestyle interventions entailing dietary and physical activity modification have been proven to prevent T2D in randomised controlled trials. Among those with

impaired glucose tolerance, the incidence of T2D was halved in the dietary intervention group (e.g. populations in the United States of America, Finland and China),[36–38] and in some populations these benefits were sustained up to 20 years post-intervention.[37] The US Diabetes Prevention program (DPP), where 3,234 high-risk participants were randomised to either placebo, Metformin or lifestyle intervention (including reduction in fat and energy intake), demonstrated that lifestyle intervention reduced the incidence of T2D by 58% after 2.8 years of follow-up and was more effective than Metformin (31% risk reduction).[39] Although impressive and clinically significant, from these trials it is difficult to distinguish the relative contribution of benefit from diet versus physical activity, which is important in understanding the mechanism to enable fine-tuning and more effective and cost-efficient planning of future interventions. It is also not possible to tease out the underlying dietary factor/s involved.

Dietary modification has been a public health priority in preventing T2D.[19,40,41] However, the optimal dietary profile is still a mystery. What is known is that for those who are overweight or obese, a dietary intake which achieves weight loss is effective in preventing T2D. Next, a brief summary of the evidence for diet in T2D prevention will be presented, at different 'levels' of the dietary exposure spectrum, from dietary patterns to foods and beverages to nutrients. Macronutrient intake was the primary dietary exposure of interest within this thesis.

### *2.5.1 Dietary patterns*

Several patterns of eating, underpinned by specific combinations of dietary components, have been demonstrated to reduce the risk of developing T2D. This includes the Mediterranean dietary pattern, the Alternate Healthy Eating Index, and the Dietary Approaches to Stop Hypertension (DASH) diet, to name a few for the prevention of T2D.[42] The longest running randomised controlled trial investigating the Mediterranean dietary pattern (median follow-up of 4.8years, N=3,541) found that compared to controls (advice on a low-fat diet), those consuming a Mediterranean dietary pattern with extra-virgin olive oil showed 40% lower risk (95%CI: 15-57%) of developing T2D.[43] Although understudied,

the benefits of these dietary patterns must be in some way also influenced by their methods of preparation as well as how, when and where they are consumed.

### *2.5.2 Foods and beverages*

A wealth of evidence from both intervention and observational studies highlights a range of foods and beverages associated with T2D, which shows commonalities with the aforementioned dietary patterns, when disaggregated.[19,42] Table 2-2 lists these according to their reported association with T2D from recently published meta-analyses.

**Table 2-2: Foods and beverages associated with Type 2 Diabetes**

| Association with T2D | Foods and beverages (reference to published meta-analysis) |
|---|---|
| Inversely associated | Fruits[44] |
| | Green leafy vegetables[44] |
| | Root vegetables[44] |
| | Wholegrain bread and cereals[45] |
| | Nuts and seeds[46] |
| | Legumes[46] |
| | Fermented dairy[47,48] |
| | Coffee[49] |
| | Tea[50] |
| Association inconsistent or null | Fish[51] |
| | Egg and egg products[52] |
| Positively associated | Sugar Sweetened Beverages (SSB)[53] |
| | Rice[54] |
| | Unprocessed red meat[55] |
| | Processed meat[55] |

There are reasons for research investigating dietary exposures at different 'levels' of the dietary exposure spectrum (i.e. from nutrients to foods to diets). Figure 2-5 illustrates this spectrum using carbohydrates. Certainly, research that provides messages about foods and dietary patterns enables improved public understanding and easier uptake of dietary advice.[19] Another reason relates to potential synergism between nutrients and non-nutritive constituents which would suggest that the sum of the whole food and/or beverage may act independently to the individual nutrients it is composed of.[19,56] The same applies

to dietary patterns. Nutrients are not consumed in isolation, however, understanding disease aetiology and biological mechanisms underpinning the association between higher level dietary factors and T2D requires examination at the level of nutrients. At the molecular level, catabolised nutrients such as glucose, amino acids, fatty acids and micronutrients may directly interact with transcription factors, ligands, DNA, etc. For example, the vitamin D receptor (VDR) is a transcription factor that triggers a cascade of cellular signalling post-DNA-binding, which controls serum calcium levels, metabolism, cellular growth and immune functions.[57] Like a lock and key, the nutrient $1\alpha,25(OH)_2D_3$ (vitamin D) needs to bind to the VDR to initiate this process. Therefore, there is merit in researching nutrients to help understand the molecular aetiology of T2D, as in the case of gene-diet interactions which will be discussed in section 2.6. Next, a summary of the latest literature regarding energy and macronutrient intake in relation to the risk of developing T2D, will be provided. The majority of published literature with T2D as an outcome is observational in nature because of the longer follow-up required in identifying the development of T2D. Other factors that may make it challenging to study dietary factors using an RCT design include motivation for long term adherence to a dietary intervention, drop outs, difficulty in blinding and cost which may contribute to bias. Since adiposity and glucose homeostasis are likely mediators between diet and T2D the literature on RCTs pertaining to these intermediate traits were also examined to gain a more complete picture of the impact of diet on T2D development.

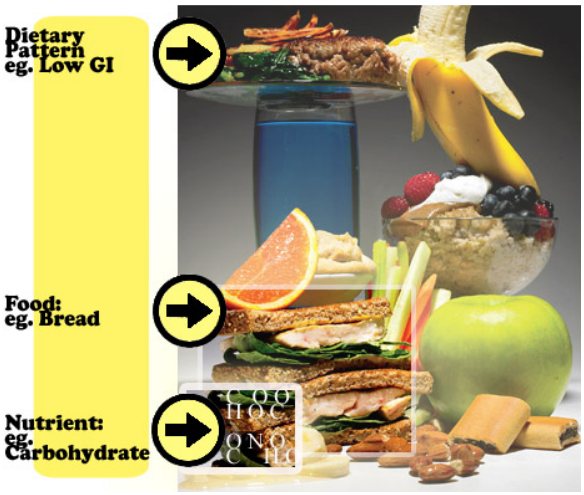



**Figure 2-5: Nutrition spectrum (designed by SL using a photo from bicycling.com)**

### *2.5.3 Energy intake affects weight whereas macronutrient composition affects glucose homeostasis*

It has been conventional clinical practice to prevent excessive weight gain by managing energy balance (i.e. energy intake *vs* energy expenditure). Energy reduction, regardless of macronutrient composition leads to weight loss, as exemplified in a meta-analysis of 48 weight loss interventions.[58] However, macronutrients have been shown to differentially influence T2D risk via their effects on glucose homeostasis and insulin resistance, in addition to how their palatability may influence adherence to the diet. Figure 2-6 summarises the observational literature, whereas Table 2-3 summarises the evidence from RCTs.

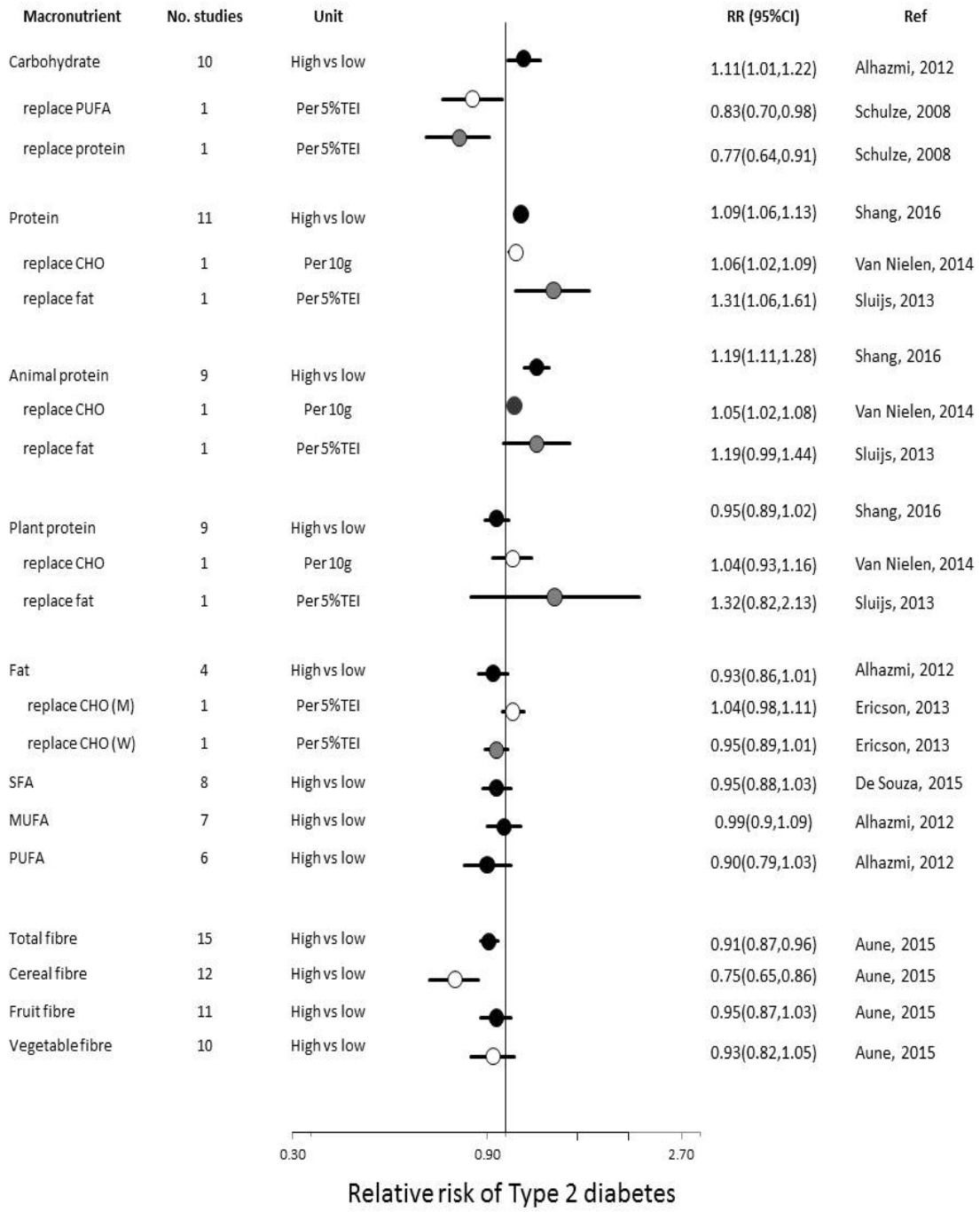



**Figure 2-6: Association between macronutrient intake and Type 2 Diabetes (observational literature).**

This figure was compiled after reviewing the literature on this topic (non-systematic). Where possible, the most recently reported meta-analysis has been reported and the estimates from the most adjusted models have been used. Example of interpretation: the risk of developing T2D is 23% lower when replacing 5% total energy intake from protein (decrease intake) with carbohydrate intake (increase intake) (Schulze et al., 2008). 'Replace' means, to reduce this macronutrient. CHO- carbohydrate, SFA- saturated fat, MUFA- monounsaturated fat, PUFA- polyunsaturated fat, TEI: total energy intake, M- men, W- women.

#### *2.5.3.1 Macronutrient quantity*

Macronutrient quantity can be measured as a proportion of total energy intake (i.e. % of total energy intake: TEI).

**Total fat intake**

Among all the large diabetes prevention trials, which involved an intensive lifestyle intervention consisting of exercise and dietary behavioural change, from my knowledge, only the Finnish Diabetes Prevention Study investigated macronutrient intake and the risk of developing T2D.[59] In 500 participants at high risk for T2D (overweight or obese with impaired glucose tolerance), those reported to be consuming >37% TEI from total fat had a higher risk of developing T2D than those with <30%TEI from total fat (HR: 2.14, 95%CI: 1.16, 3.92).

Within meta-analyses of clinical trials for intermediate traits there is inconsistency in the effect of total fat intake on weight change, among overweight, obese and/or normal weight individuals (for weight maintenance).[60–62] Whilst Hooper et al., noted an overall weight reduction in favour of lower fat diets in a meta-analysis of 33 RCTs (every 1% decrease in energy from total fat led to 0.19kg decrease in body weight compared to no change in fat intake),[61] a more recent meta-analysis of 53 RCTs concluded no statistically significant difference in weight loss between low fat and higher fat dietary interventions of similar intensity (i.e. time and resources or support provided by study staff) (weighted mean difference in weight loss: 0.36, 95%CI: -0.66 to 1.37).[62] Tobias and colleagues suggest that the difference observed with Hooper et al., may relate to confounding by the intensity of dietary interventions among studies of longer than one year in duration. Weight loss trials have shown that adherence to allocated diets diminishes with increasing duration, so it is unclear to what extent this influenced the reported findings. Tobias et al., used triglyceride levels at follow up as a proxy for adherence to a low fat diet,[62] however, it is questionable how well this captures adherence to a low fat diet, given that exercise, alcohol intake and medications can also influence triglyceride levels.

Observational cohort studies do not support a major influence of total dietary fat on the risk of T2D[63] but allude to the differential effects of fat subtypes (i.e. SFA, MUFA, PUFA), which will be discussed under the section titled 'macronutrient quality'.[42]

### Total carbohydrate intake

Low carbohydrate diets have been popular but also controversial. The definition of a 'low carbohydrate diet,' varies, with studies including diets from 20g/d such as the Atkins diet[64] to <45% of total energy from carbohydrate intake,[65] and advocating restriction or avoidance of certain foods. For example, the South Beach diet is based on the philosophy of training your body to use fat as the main source of energy, starting off with avoidance of sugar and refined carbohydrates as well as fruits and whole grains.[66] The Banting diet recommends avoidance of gluten containing grain products.[67] To my knowledge, there is no published systematic review of clinical trials investigating the effect of carbohydrate intake on the risk of developing T2D.

Two meta-analyses demonstrated that there is little difference in the change in blood glucose levels between those who consumed low carbohydrate (<45% energy) compared to low fat diets (<30% energy).[64,65]

Two meta-analyses have shown that low carbohydrate diets results in greater weight loss compared to low fat diets.[62,64] For example, Tobias and colleagues found that among 18 RCTs, those on a low carbohydrate diet experienced a larger weight reduction than those on a low fat diet (weighted mean difference of 1.15kg more in weight reduction, 95%CI:0.52-1.79).[62] Whereas other meta-analyses have shown little difference in weight loss between a low carbohydrate diet compared to either a low fat diet or those with a moderate macronutrient composition.[58,65,68] A major caveat of studies that compare low carbohydrate and low fat diets is that, under isocaloric conditions, low carbohydrate diets may also represent an increase in fat or protein intake or an equal increase in both. Therefore, the independent effect of reducing carbohydrate intake cannot be deduced from these studies. Moreover, low carbohydrate diets have been shown to increase LDL

Cholesterol,[64,65,68] yet also favourably increase HDL Cholesterol and decrease triglyceride levels.[64,65,69] However, it is unclear whether there are other non-metabolic adverse consequences with low carbohydrate diets and at what level of carbohydrate intake would they manifest.

In the observational setting, total carbohydrate intake is not associated with T2D.[63] However, in over 27,000 Germans, there was a suggestion of an inverse association when 5%TEI from protein was modelled to be replaced by carbohydrate intake (HR: 0.77, 95%CI: 0.64, 0.91).[70] It is unclear, however, to what extent this observed association was related to carbohydrate quality such as whether the carbohydrate was from whole- or refined-grain products.

### Total protein intake

The definition of a, 'high protein diet,' varies but typically is >20% energy from protein intake. It has been purported to increase body fat mass loss, attenuate fat free mass loss, increase satiety and improve cardiometabolic profile. Among short term RCTs (<6 months duration), the effect of high protein diets on plasma insulin in healthy individuals is inconclusive whereas amongst those who are overweight or obese, high protein diets do not appear to improve insulin sensitivity over and above the effect of weight loss.[71]

For weight loss, high protein diets appear to be beneficial, at least in the short term.[72] Santesso and colleagues found that when 38 trials comparing high (ranging from 16-45% of total energy intake) to low protein diets (ranging from 2-23% of total energy intake) were meta-analysed, there was a statistically significant larger weight loss in those consuming high protein diets (SMD of 0.36kg, 95%CI: 0.17, 0.56 greater), albeit high heterogeneity was also evident ($I^2$=77%).[72] Other measures of adiposity, including BMI and waist circumference showed directionally similar effects in favour of a high protein diet, at three months.[72] However, the benefits attenuated when examining findings for studies of longer than three months in duration.[72] This was corroborated by the findings from two other meta-analysis, one comparing high with standard protein diets (N=13 studies)[73] and another comparing high with low protein diets (N=13 studies).[74] It appears that the benefits of high protein diets are temporary, however, it is unclear whether

this benefit may be confounded by the simultaneous reduction in glycaemic load whilst consuming a high protein diet. Similarly with low carbohydrate diets, there is improvement in HDL-Cholesterol and reduction in triglycerides but some have reported adverse gastrointestinal symptoms including bloating, constipation and diarrhoea.[72]

Another systematic review showed that among prospective observation studies, those who consumed the highest amount of total protein had a 9% higher risk of developing T2D (95%CI: 6, 13%), compared to those who consume the lowest.[75] When modelling replacement of total protein for total fat or carbohydrate intake, findings did not materially change.[76,77]

#### *2.5.3.2 Macronutrient quality*

**Fat subtypes**

A meta-analysis of clinical trials concluded that glycaemic indices, including plasma glucose, HbA1c and HOMA-IR improved when unsaturated fats (either monounsaturated or polyunsaturated fat) were consumed instead of saturated fat or carbohydrate.[78] This confirms the importance of fat subtypes in glucose homeostasis and subsequent T2D propensity. It also highlights that the replacement fat is as crucial as the type of fat being replaced. Two meta-analyses, one of 20 cohort studies (RR: 1.03, 95%CI: 0.98, 1.07)[79] followed by another of 12 observational studies (RR: 1.06, 95%CI: 0.95, 1.17)[80] both showed that SFA was not associated with coronary disease risk. However, their inferences suffered criticism after evidence was brought forward about differential associations between SFA and CHD, depending on the macronutrient SFA was being replaced by, which may mask an association when looking at SFA by itself. For example, a meta-analysis of eight RCTs found that the overall risk of CHD was reduced by 10% (95%CI: 3-17%) when 5%TEI from SFA was replaced by PUFA.[81] Similarly, observational studies demonstrated similar benefits, of a reduction in CHD risk, when SFA was replaced with PUFA (HR: 0.75, 95%CI: 0.67, 0.84), MUFA or complex carbohydrates.[82] Whereas, CHD risk was not associated with SFA when it was replaced with intake of simple carbohydrates.[82] This differential association was evident because the researchers modelled isocaloric

macronutrient substitution, which will be explained in more detail in Chapter 3, and highlights the importance of this method as well as macronutrient quality when investigating macronutrient intake and disease relationships.

**Table 2-3: Association between macronutrient intake and Type 2 Diabetes (from RCTs)**

| Direction of association | Outcome | Macronutrient (reference to published meta-analysis) |
|---|---|---|
| Inversely associated | Glycaemia and IR | MUFA replacing SFA[83]<br>PUFA replace SFA[83]<br>PUFA replace CHO[83] |
| | Obesity | Total protein (short term <3months)[72–74]<br>Total protein replace CHO (but increases mortality and cardiovascular disease risk)[60,68] |
| Association inconsistent or null | T2D<br>Glycaemia and IR | SFA[59]<br>SFA replacing CHO[83]<br>PUFA replace MUFA[83]<br>Total CHO[64,65] |
| | Obesity | Total fat [62]<br>Total CHO[60,65,68]<br>Total Protein[60]<br>Different macronutrient compositions show little difference between each other[58] |
| Positively associated | T2D<br>Obesity | Total fat[59]<br>Total fat[60,61]<br>Total CHO (low carbohydrate diet increase weight loss)[62,64] |

Abbreviations: RCT- randomised controlled trial, T2D - Type 2 Diabetes, IR- insulin resistance, MUFA- monounsaturated fatty acids, SFA- saturated fatty acids, PUFA- polyunsaturated fatty acids, CHO- carbohydrate

### Animal versus plant protein

A meta-analysis of 13 RCTs reported glycaemic benefits (HbA1c, fasting glucose and insulin) of replacing animal protein with plant protein.[84] However, these studies were among participants with established diabetes, some on medication, and 85% of the studies were of poor quality and of <12 weeks in duration. A meta-analysis of 11 cohort studies also find that animal protein is positively associated with T2D (RR: 1.19, 95%CI: 1.11, 1.28) whereas a null association was observed for plant protein intake (RR: 0.95, 95%CI: 0.89, 1.02).[75] In the Nurses' Health Study and Health Professionals Follow-

up Study, researchers found that modelling replacement of 5% energy from animal with plant protein intake lowered T2D risk by 23% (95%CI: 16-30%).[85] They also revealed that beyond the classification of animal and plant protein, the source of animal protein that is being replaced can significantly alter this relationship. That is, in a meal with refined grains and potatoes, when a serving of processed red meat was modelled to be replaced by a serving of plant protein-based food (wholegrains, legumes, nuts) this resulted in a 21% lower risk of developing T2D (95%CI: 16-26%), whereas this effect was non-significant when modelled to replace a serving of dairy (HR: 0.98, 95%CI: 0.96, 1.00).[85] These suggest that protein quality and food source differentially influences T2D risk.

#### Dietary fibre

Evidence from a meta-analysis of cohort studies reported that consuming more dietary fibre, in particular, cereal fibre was associated with a lower risk of T2D (RR of T2D comparing high vs low consumers of cereal fibre: 0.75, 95%CI: 0.65,0.86).[86] This included results from the EPIC-InterAct study (HR: 0.81, 95%CI: 0.70, 0.93, which was attenuated after adjusting for BMI).[86] Similarly, clinical trials providing grains such as rye, barley, oats and fibre supplements including psyllium (insoluble fibre) have noted clear benefits to acute postprandial blood glucose and insulin as well as long-term insulin sensitivity.[87] A meta-analysis investigating the isolated effect of psyllium among euglycaemic individuals in clinical trials ranging from 2 to 26 weeks discovered a directionally consistent lower postprandial blood glucose (-12.4 mg/dL, 95%CI: -17.5, -7.3, 11 publications) and insulin (-126.8 mg/dL, 95%CI: -219.5, -34.1, 6 publications) compared to placebo.[88] There is less consistency in the effect of dietary fibre for preventing obesity, with hypothesised mechanisms related to satiety and appetite still unclear.[87] There is also a lack of association between fibre intake from fruit and vegetable sources and T2D.[86,89]

In summary, evidence for macronutrient intake and T2D as well as for intermediate traits is clearly complex. Although it is reasonably clear that total energy intake affects weight, macronutrient intake and more importantly which

substitutions appear to differentially drive glucose homeostasis to the detriment of T2D development remains to be fully elucidated. The current trial evidence for total macronutrient composition with body weight and glucose homeostasis is inconsistent. Notably, trial data show that the quality indicators of macronutrient intake, including PUFA, plant protein and cereal fibre intake appear to improve glycaemia. Whereas, observational studies investigating the risk of T2D show inconsistent associations with macronutrient composition, except for total and animal protein and dietary fibre. From the literature, three key issues were highlighted:

1. A key limitation that prevents clear comparison between trials comparing diets with different macronutrient compositions is that many of these trials report insufficient information about the relative proportions of energy from each macronutrient (e.g. does a low-fat diet represent a diet high carbohydrate and protein or only higher in protein?). Therefore, this makes clinical interpretation difficult.
2. Diets which were very restrictive are unlikely to be sustainable in the long term because of the difficulty in adherence. Adherence is currently poorly measured in most trials using *ad-libitum* diets. Additionally, few studies report adverse consequences of dietary interventions, but this is important in helping to identify thresholds for macronutrients intake (both minimum and maximum) to minimise health risk beyond deficiency in essential nutrients (e.g. essential amino acids or fatty acids). For example, a low carbohydrate diet which may also be high in protein is likely to be low in dietary fibre and may increase the likelihood of renal complications or electrolyte imbalance but this remains to be studied.[90] Indeed, a study which analysed such diets (e.g. Atkins), albeit only over three days, discovered that six micronutrients (e.g. vitamin B7, E and iodine) were either insufficient in meeting nutrient requirements or non-existent.[91] This raises concerns about potential long term micronutrient deficiencies.
3. Macronutrient quality and food source.

The inconsistencies in response to macronutrient intake on the risk of developing T2D between studies beg the question about whether biological differences such as those from genetic variants may play a role. That is, can gene-regulated

mechanisms of digestion, absorption, metabolism and distribution of macronutrients affect nutrient bioavailability and therefore downstream effects on glucose control and T2D risk differently between study participants and between studies? There is emerging research that tries to evaluate this within a field of study known as gene-diet interactions.

## *2.6 The interplay between genetics and diet in the prevention of type 2 diabetes*

### *2.6.1 Definitions*

The notion that a multifactorial condition such as T2D is a likely result of the interplay between our genes and environment is well accepted but the exact mechanism is unknown. Within the broad scope of research that examines the interplay between genes and environment I focused my investigation on gene-environment interactions with a particular lens on gene-diet interactions. In some fields of research this overlaps with 'nutritional genomics' (see Box 1), specifically 'nutrigenetics,' which examines how genetic variation influences the effect of diet on disease or on nutrient requirements. This is distinct from 'nutrigenomics.'[92] Within this thesis, I will use the term 'gene-diet interactions' or a variant of this. More details relating to the statistical concepts of interaction will be provided in Chapter 3.

**Box1: Definitions**

**Nutritional genomics** is an umbrella term, encompassing both nutrigenetics and nutrigenomics.
**Nutrigenetics** investigates how genetic variation affects nutrient requirements as well as the interaction with nutrients on disease development. This may be due to genetic polymorphisms encoding enzymes or transporters involved in nutrient ingestion, digestion, absorption, metabolism and/or elimination.
**Nutrigenomics** examines how nutrients may alter the genome, proteome and metabolome.[92]
Nutrigenomic studies may contribute to discussion of my findings as they aid in the broader aetiological understanding of nutrient-gene interactions but will not be specifically examined as this is out of the scope of my PhD.

### *2.6.2 Motivation for studying gene-diet interactions*

There are two primary reasons motivating research in gene-diet interactions within T2D:

1. Improve the understanding of T2D aetiology:
    a. Investigating gene-diet interactions, alongside epistasis (gene-gene interactions) may help to explain the 'missing heritability' of T2D.[35]
    b. Help to identify novel genetic variants without any marginal effects that may be masked by environmental exposures such as dietary intake. Conversely, this can also help identify novel environmental factors that are only evident among those with particular genetic susceptibility.[35]
    c. Statistically, modelling the combined contribution of genes, environment and their interplay on T2D development, will more accurately explain the effect of each of these factors.[93]

2. Improve the method of identifying those at risk of T2D and therefore better prevent it: a classic example of gene-diet interaction is phenylketonuria (PKU), which only manifests in those who consume phenylalanine whilst possessing a recessive mutation in the gene phenylalanine hydroxylase. Children with this condition avoid neurological damage when adhering to well-established dietary interventions based on knowledge of this interaction.[94] This is a monogenic condition. Since then, advancements in genetic technology and the sequencing of the human genome are enabling gene-diet interaction research on polygenic and multifactorial disorders, such as T2D. This may aid in identifying individuals in a population who respond well to dietary interventions that prevent T2D, those who do not (non-responders) as well as those who experience worse outcomes (adverse responders), assuming similar levels of adherence. Although appealing, whether this may translate into genetically stratified approaches for optimising identification and provision of more effective dietary interventions in preventing T2D, is yet unknown.[35,92,95]

This vision for translation, for which currently there are no concrete and experimentally validated examples for T2D, [96] has been coined by some as 'personalised' or 'precision' nutrition.[97–99] It has gained much attention, often secondary to a parallel field called personalised medicine, which in recent times both the UK and US government have invested large amounts of public funding to its research.[100,101] Within the dietetics profession, personalised nutrition was identified as a 'change driver' by the Dietetics Workforce Supply and Demand Future Scan.[102]

Some of the public health benefits envisaged from genetics based personalised nutrition include:

1. Refinement of targeting dietary interventions to people most likely to respond well to them;
2. Improvement in cost and health effectiveness of existing preventive strategies and treatments related to diet (e.g. medical nutrition therapy);
3. Minimising adverse effects related to dietary interventions;
4. Improvement in patient adherence to clinical recommendations; and
5. Discovery of novel dietary interventions that are beneficial only in a defined genetic subgroup of the population.[95]

How these may be implemented into practice is under considerable discussion and research. This includes whether it is administered in person with a health professional or online (e.g. the Food4Me Study); how genetic risk itself should be communicated to optimise comprehension and behaviour change; can it operate safely, ethically and equitably through private health services or does it need to be via public health services; and what level of scientific evidence is sufficient for translation? Real life examples from commercial companies are provided in table 2.5.

### *2.6.3 Examples of gene-environment interactions in multifactorial conditions*

Cancer epidemiology has elucidated some of the most replicated gene-environment interactions to date and is considered a leader in genetic-personalisation of diagnosis and treatment. For example, given that not everyone

who smokes tobacco develops cancer, it was hypothesised that certain polymorphisms such as that in the *NAT2* gene (rapid or intermediate vs slow acetylators) which has a role in detoxification of carcinogens may modify cancer risk from tobacco smoking. Indeed, a case-only meta-analysis (n=22 studies) found that smokers who were genetically predisposed to slow acetylation had a significantly higher risk of bladder cancer compared to those who were intermediate or rapid acetylators (interaction OR: 1.2, 95%CI: 1.1-1.5, $p_{interaction}$=0.008).[103] In psychiatry, examples such as whether a polymorphism in the *5-HT T* genetic locus differentially influences how stressful life events and childhood maltreatment predicts depression has caused much controversy.[104] This interaction has had one of the highest attempted replications, with mixed results, so conclusions are still under intense debate and scrutiny.[105]

Within the cardoiometabolic literature, a meta-analysis of >200,000 European and North American adults investigated one of the most replicated interactions.[106] Authors found that the effect of the strongest obesity-associated locus (*FTO*) on obesity risk was attenuated among those who were physically active by 0.4kg/minor allele when compared to those who were sedentary. This was also observed within UK Biobank (N~ 120,000),[107] and the Food4Me study which used objective measures of physical activity (n=1,280).[108] However, an individual participant meta-analysis of 8 RCTs (n=9,563) which had investigated either dietary, drug or lifestyle intervention (diet and physical activity) on body weight, did not observe differences in response when individuals were stratified by *FTO* genotype.[109] The discrepancy between observational and trial data, in this case, may be explained by i) possible measurement bias in self-reported physical activity and the cross-sectional nature of some observational studies (e.g. UK Biobank) and ii) the trials did not study the independent effects of physical activity and may be underpowered to examine this interaction.

Several dietary interventions with varying macronutrient composition also suggest a genetic contribution to variation in weight loss response.[60] Similar genotype-specific responses to dietary interventions have also been reported for insulin resistance,[110] and T2D risk,[111] though not all findings have been consistent.[112,113] For example, in a weight loss trial of 737 overweight adults, Qi et al., observed a significantly greater improvement in glucose homeostasis (fasting

glucose, insulin and HOMA-IR) among those possessing the TT genotype of rs2287019 (*GIPR* locus) and consuming a low fat diet compared to a high fat diet at 6 months follow-up.[114] Whereas for those with either the CC or CT genotype, the choice between low and high-fat diets did not matter.[114] Additionally, whether individuals with the genetic variant in the *APOE* locus, that are predisposed to coronary heart disease (E4+), are more sensitive to the lipid raising effects of dietary fat intake have also been of interest. In an adequately powered feeding trial, with participants recruited based on APOE genotype (N=88, duration=8 weeks), authors reported a significantly greater decrease in circulating triglyceride levels in *APOE4* carriers compared to non-carriers following a low fat intervention (0.48±0.11mmol/L and 0.22±0.06mmol/L, respectively).[115] Whereas this response was not observed for changes in total cholesterol in a large multi-centre European *ad-libitum* dietary intervention trial (N=1,466, Food4Me, follow-up=6months).[116] This begs the question of, if genetic variation establishes an individual's baseline risk for developing T2D could diet, a known modifiable modulator of this risk, accentuate, attenuate or even abolish T2D risk?[117] This instigated further questions about whether there are indeed replicated, validated and clinically useful interactions that may challenge our current one-size-fits-all nutritional recommendation for the prevention of T2D.

### *2.6.4 Current methodological challenges facing the field of gene-diet interactions*

Despite these intriguing research questions, several reviews on gene-lifestyle interactions for cardiometabolic conditions do not highlight any specific interactions that have been replicated or with a magnitude that is clinically relevant to take forward for application.[96,118] Several reviews also present the methodological challenges that make researching gene-environment interactions difficult and these challenges may help to explain the current state of evidence.[95,105,119,120] These are summarised in Table 2-4, including a brief introduction to the approaches taken within this PhD to overcome some of these.

| Table 2-4: Key limitations of current gene-diet interaction studies on T2D and methods that have been proposed to overcome these | | |
|---|---|---|
| **Key challenges** | **Explanation** | **What have I done to address this** |
| Sample size | In general, common SNPs have modest effects on multifactorial outcomes (on average below 20% higher risk). Therefore, large studies, often in the tens of thousands of individuals are needed to detect these effects. Subsequently, the number of people needed to study interactions between genetic variants and diet would be even larger.[120,121] This also depends on the minor allele frequency of the genetic variant of interest. | Increasing sample size, using a large prospective cohort with incident T2D cases[119] |
| Publication bias | Often novel, interesting gene-diet interactions are published without replication of findings. Conversely, negative results are often unpublished. | Attempt to publish all interaction analyses, regardless of findings. |
| Measurement bias of dietary exposures | How accurately we can estimate a population's intake using self-reported methods e.g. questionnaires are a well-known problem in nutritional research. Participant burden is balanced against the precision of measurement.[120,122] | Sample sizes could be reduced by 30 fold with increased precision of exposure measurement, such as by using objectively measured data.[122] |
| Possible selection of inappropriate genetic targets | i) simple interaction analyses based on single candidate variants may not reflect the complex biology of living systems nor the polygenic nature of T2D, where epistasis may exist,[123] ii) genetic variants without demonstrable marginal effects on T2D may exhibit synergistic interaction under environmental influence.[124] | Genetic risk scores (GRS) has enhanced power in detecting interactions.[35,125] Also, hypothesis-free based genome-wide interaction analysis (GEWIS) will aid in discovering novel interactions at potentially unexpected genetic loci.[35] |
| Confounding | In studies of interaction, both sets of confounders for each exposure are equally important to account for. Few studies have accounted for population stratification and confounding relating to dietary exposures, which may lead to spurious findings.[120,126,127] Nor have they addressed potential confounding between genetic or dietary exposures and covariates.[128] | Adjustments for population stratification and other key confounders. |
| Multi-collinearity | Multicollinearity between exposures and their interaction variables can lead to inflated estimated variances of the interaction coefficient. | A simple method of mean centring of continuous exposure variables has been shown to reduce covariance and correlations between main exposures and the interaction term.[129] |
| Dynamics of interactions | Exposures and therefore interactions may change over the lifetime of an individual and population.[120] | Beyond the scope of this PhD |
| Limited reproducibility | Influenced partly by the above factors, many reported gene-diet interactions have failed to be replicated in similar populations. This questions the validity of such interactions. | Attempt to replicate reported interactions available in the current literature but also conduct replication if I identify any statistically significant interactions. |

### *2.6.5 Current societal and commercial backdrop*

With genotyping and genome sequencing becoming faster and more affordable, a number of commercial companies are providing direct-to-consumer lifestyle tests and dietary advice, predicated on the evidence of gene-diet interactions, and have been under immense scrutiny by regulatory agencies.[130,131] Some examples have been listed in table 2-5. Although conflicting findings have been reported for its clinical utility (research examining whether genetic risk information and/or based dietary advice leads to change in dietary behaviours) and validity (research examining gene-diet interactions), particularly in the area of T2D prevention, the overall evidence taking into account the quality of studies is yet to be determined. Whereas the ethical, social and legal issues of these 'personalised nutrition' services, including whether genetic test results induce psychological harm, has also been called into question and requires further investigation.[132] Therefore, investigating gene-diet interactions is urgently needed to confirm or refute unambiguously the current clinical validity of genetic-based personalised nutrition. Findings may provide timely evidence and advice for healthcare professionals and policy makers to better guide and safeguard the public.

| Table 2-5: Examples of commercial organisations providing 'personalised nutrition' advice | | | |
|---|---|---|---|
| **Company name, Headquarters and founding year** | **Tests offered for the following, Number of genes this is based on** | **More information, including costs and their values on science** | **Website** |
| *Nutrigenomix*<br>Canadian<br>Founded in N/A | Cardiometabolic health<br>Nutrient metabolism<br>Weight management<br>Food intolerances<br>Eating habits<br>Physical activity<br>Injury risk<br>Based on 45 genetic markers | DNA analysed via saliva sample<br>Operates through dietitians<br><br>Price: N/A<br><br>'Genetic tests are based on the most robust scientific evidence.' | https://www.nutrigenomix.com/ |
| *DNAnalysis*<br>South-African<br>Founded in 2007 | Obesity risk/insulin resistance: FTO<br>Binge eating: DRD2<br>Carbohydrates and obesity: ADRB2<br>Fat absorption: FABP2<br>And other genes not listed on their website | Operates through a health care professional<br><br>Equivalent to £100<br><br>'DNA Diet is designed to assist the healthcare practitioner in the design of a personalised healthy eating plan based on individual genetic differences' | http://dnalysis.co.za/dna-diet/ |
| *DNAFit*<br>UK<br>Founded in 2013 | Diet type recommendation<br>Carbohydrate response<br>Saturated fat response<br>Lactose intolerance<br>12 week recipe guide<br>Genotype specific shopping list<br>Detoxification genetics<br>Anti-oxidant needs<br>Omega-3 need<br>Vitamin B need<br>Vitamin D need<br>Alcohol sensitivity<br>Caffeine sensitivity<br>Salt sensitivity<br>20 genes listed on their website | DNA analysed via saliva sample<br>Direct to consumer<br><br>£99-199<br><br>'(genes included)... in the DNAFit reports only once the evidence is clear regarding an easily modifiable gene x environment interaction with a positive outcome, and that this effect is shown in multiple studies.' | https://www.dnafit.com/diet |

| | | | |
|---|---|---|---|
| *Vitagene*<br>USA<br>Founded in N/A | Obesity risk<br>Alcohol metabolism<br>Cholesterol levels<br>Triglyceride levels<br>Lactose sensitivity<br>Gluten sensitivity<br>Emotional eating<br>Weight regain after dieting<br>Fat intake<br>Sodium intake<br>Supplementation test: which supplements to take | Supplemental service: uses genetic results from other genetic testing companies (e.g. 23andMe) to provide a secondary analysis service that advises on dietary intake.<br><br>Equivalent to £38 | https://vitagene.com/products/ |
| Personalised eating restaurant<br><br>*VITA MOJO*<br>UK<br>Founded in 2016 | VITA MOJO and DNAFit teamed up to provide menus created based on 'personalised nutrition,' advice provided by genetic testing from DNAFit.<br><br>Website allows you to adjust how much of the dish you choose to purchase, to match the macronutrient composition that is presumably recommended by your DNAFit results. | | https://www.vitamojo.com |

## *2.7 PhD framework, aims and objectives*

The current literature demonstrates strong epidemiological evidence for the individual effects of genetics and of diet in the development of Type 2 Diabetes (T2D). Theoretically, gene-environment, including gene-diet, interactions plays a role in T2D development, however, few robustly replicated empirical examples exist for cardiometabolic diseases, including T2D. There is a need to understand if and how gene-diet interactions contribute to T2D risk. This is set against the background of rising public and commercial interest in translating gene-diet interactions into 'personalised nutrition' advice. It is therefore timely to systematically evaluate the current evidence for gene-diet interactions in T2D aetiology, particularly focusing on macronutrient intake. Moreover, there are many methodological challenges currently hindering our epidemiological understanding of whether reported gene-diet interactions are in fact real. Therefore, by addressing some of these methodological challenges and by systematically and comprehensively assessing gene-diet interactions using a battery of methods, it is hoped that the aim and objectives below could be achieved. This is illustrated in Figure 2-7. The methods used within this PhD spanned the spectrum of methods for studying gene-diet interactions, from hypothesis driven genetic variant selection (biological candidate genes) to hypothesis-driven analysis based on hypothesis-free selection of genetic variants (genetic risk scores using BMI, IR and T2D associated loci) to completely hypothesis-free analysis of interaction using genome-environment-wide-interaction-study (GEWIS) that examines interaction without restriction on the genetic exposure.

**Aim:**

> Investigate the contribution of gene-diet interactions to the aetiology of T2D. Indirectly this may help evaluate the potential for using this insight to identify population subgroups that differ significantly in their response to dietary interventions.

**Objectives:**

1. Conduct a systematic review of gene-macronutrient interactions and T2D (described in Chapter 4) to understand the available literature on a specific aspect of gene-diet interactions in the development of T2D.

2. Attempt to replicate identified interaction/s in a large prospective case-cohort study (EPIC-InterAct) to evaluate the validity and reproducibility of these identified interactions (Chapter 4).

3. To examine interactions between macronutrient intake and genetic risk scores in the development of T2D (Chapter 5).

4. To examine interactions between food and beverage intake and genetic risk scores in the development of T2D (Chapter 6).

5. To examine interactions between macronutrient intake and genetic variants on T2D, across the entire genome (Chapter 7).

6. Appraise the literature as to whether there is clinical utility for providing genetic risk information on dietary behaviour change. This is described in the discussion (Chapter 8).

The 'analytic validity, clinical validity, clinical utility and associated ethical, legal and social implications' (ACCE) framework for evaluating genetic tests,[133] which share similar broader goals for public health to our work helps to place my aims and objectives into context. This framework illustrates the critical areas that are important for evaluation, which is relevant for both genetic testing and the potential use of genetic tests to personalise dietary interventions (the potential application for researching gene-diet interactions).

**Hypothesis:**

We hypothesised that based on prior literature we would detect gene-diet interactions between specific dietary macronutrients and genetic variants on the incidence of type 2 diabetes.

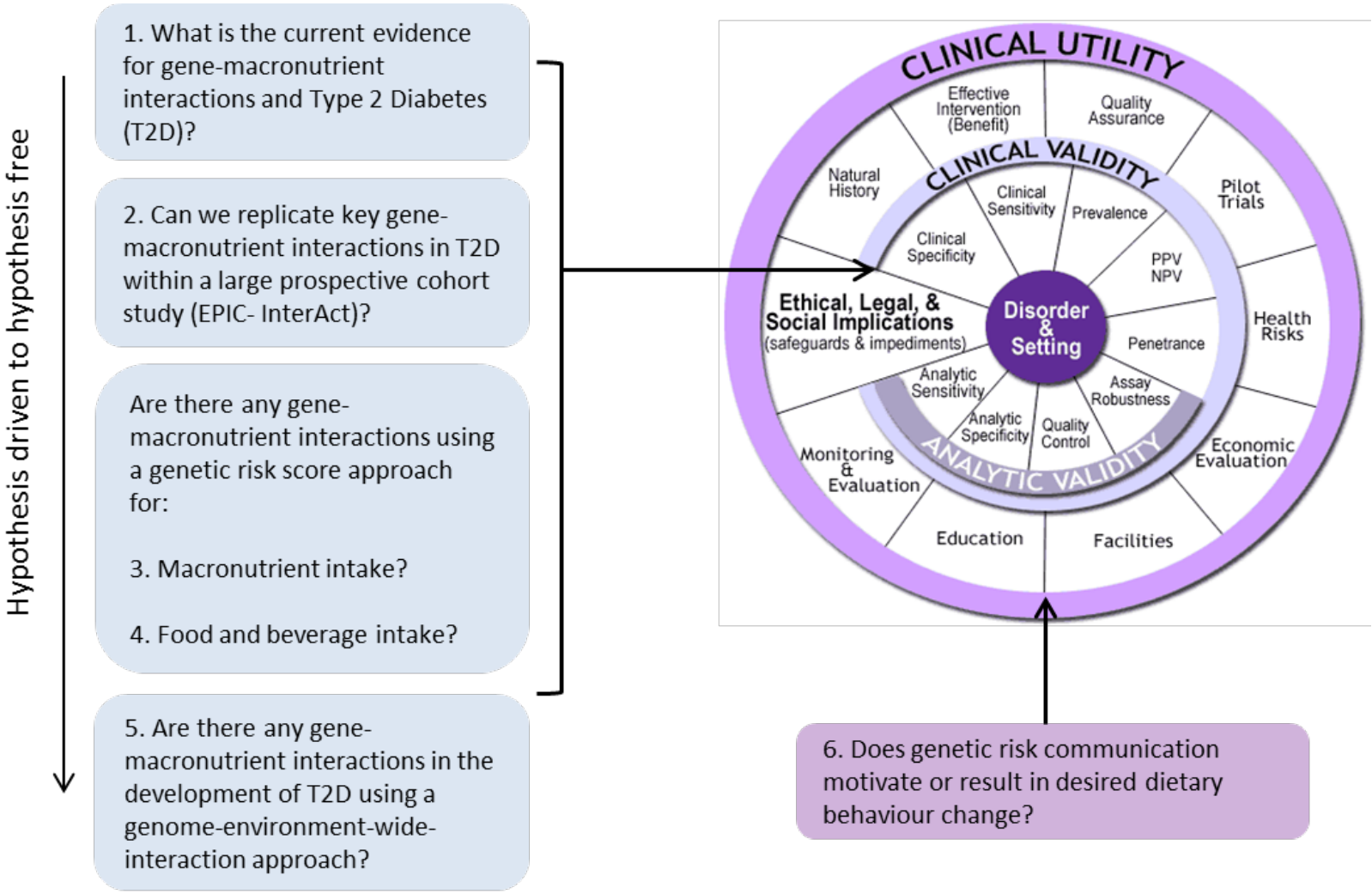



**Figure 2-7: Visual representation of the aims and objectives of my PhD, with links to the ACCE framework.[6]**

# *Chapter 3 Methods for interaction analyses and cohort descriptions*

## 3.1 Interaction analysis

### *3.1.1 What is an interaction?*

A distinction must be made between *biological and statistical interaction*, the latter of which is the focus within this thesis. Biological or mechanistic interactions often refer to a physical or chemical reaction between two compounds such as when iron inhibits gastrointestinal calcium absorption. Rothman describes it as synergism or antagonism in the 'sufficient cause framework' or 'causal pie.'[134] Synergism occurs when either two or more factors cannot independently result in an outcome but do when present simultaneously, such as when two people who have complementary skills work together to achieve a task that could not be achieved alone. Whereas, antagonism is when a risk factor must be absent for the sufficient cause to exist, for example when native bacterial flora out competes with foreign bacteria to prevent infections.[134] Statistical interactions are purely statistical. Some mechanistic interactions can be picked up statistically and others not. Conversely, sometimes even if a statistical interaction is present a mechanistic one may not be and therefore functional studies are often required to follow up on statistical interactions.[135]

In the realm of statistical interactions, the term *'effect modification'* is often confused with *'interaction'*. Attempts to clarify terminology include the distinction made by VanderWeele:[136]

**Effect modification:** the causal effect of the primary exposure (A) differs across strata of another factor of interest (exposure B). This, therefore, involves only one hypothesis and one set of confounders for the primary exposure (A). Figure 3-1.i.

**Interaction:** the causal effect requires both exposures (A and B) and therefore involves two hypotheses and two sets of confounders for each exposure. Figure 3-1.ii.

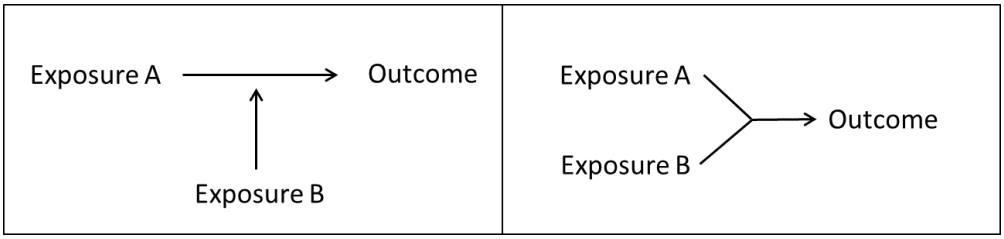



**Figure 0-1: Distinction between statistical (i) effect modification and (ii) interaction**

Within this thesis, the premise is to investigate 'interactions.' If these are found, further analyses would be conducted to distinguish whether it is purely an interaction or an effect modifier by examining strata-specific effects of each exposure one at a time.

### *3.1.2 Interaction scales*

Interactions can be estimated by introducing a product term between the two exposures in the statistical model. Below provides an example of this modelled for the interaction between G and SFA (main exposures), their interaction term (G x SFA) and covariates (C) (Equation 1).

$$f(T2DM) = \beta_o + \beta_G G + \beta_{SFA} SFA + \beta_{G \cdot SFA} G \times SFA + \sum \beta_C C_x \quad (1)$$

Statistical interactions can be assessed on one of two common scales, *multiplicative or additive*. Whilst there continues to be debate over the pros and cons and when to use either, Table 3-1 shows a summary of the differences between these two scales.

| Table 0-1: differences between the two interaction scales | | |
|---|---|---|
| **Scale** | **Multiplicative** | **Additive**[137–139] |
| Primary aim of examining interaction | Understanding of aetiology and biology. Using just the multiplicative interaction scale can indicate the wrong subgroup to treat, as demonstrated by Markowitz and colleagues.[140] | Help identify which group in the population to target finite healthcare resources to. More relevant for public health. |
| What it measures | Relative risk of different groups (relative ratio) | Difference in number of individuals at risk (absolute rate) |
| Regression model suitable | Multiplicative models e.g. Cox regression | Additive models e.g. attributable risk |
| Further explanation | Combined effect of the two exposures (i.e. genetic variants and diet) would be larger or smaller than the product of the individual effects [$HR_{GRS\ x\ diet}$/($HR_{GRS}$ x $HR_{diet}$)] <1 or >1 | Whether an absolute incidence rate due to the combined exposures (i.e. genetic variants and diet) would be larger or smaller than the sum of incidence rate due to each individual exposure [ $HR_{GRS\ x\ diet}$ – $HR_{GRS}$ – $HR_{diet}$ + 1] |
| How it is calculated | $OR_{11}$/ ($OR_{10}$ x $OR_{01}$) | $P_{11}$ - ($P_{01}$ + $P_{10}$) + $P_{00}$ (if $P_{00}$= 1) = $P_{11}$ - $P_{01}$ - $P_{10}$ + 1 |
| Strength | Computationally faster. Relative ratios are less heterogeneous than absolute rates. | |
| Caveat | Must interpret with knowledge of the main effects of each exposure. | |

Although VanderWeele recommends reporting interactions on both scales because additive interactions can exist independent of multiplicative interactions,[139] the decision on interaction scale was guided by the objective of the specific project. That is, when examining the contribution that gene-diet interactions have on the aetiology of disease, the multiplicative scale was adopted. Whereas when dietary exposures already have public health implications (e.g. sugar-sweetened beverage intake is detrimental for T2D),[53] this also justified examining additive interactions.

### *3.1.3 Methods in studying gene-diet interactions*

Approaches to examining gene-diet interaction can be broadly classified into hypothesis-driven or hypothesis-free approaches and both entail methodological limitations. Hypothesis-driven approaches may include examining i) single genetic variants, using either a candidate gene approach or GWAS-derived genetic variant/s, or ii) a combination of variants such as a GRS or haplotype.

Taking a hypothesis-free approach involves investigating possible interactions across the whole genome, where interaction is tested between an exposure and every measured or imputed genetic variant across the genome. This is known as GEWIS. [35] All these three methods have been used in this thesis. Table 3-2 provides an explanation of these methods and others, including their strengths and limitations.

### *3.1.4 Quality criteria for evaluating interactions*

Currently, there is no well-accepted published recommendation on determining the quality of reported gene-diet interactions, or of gene-environment interactions in general. A common and accepted list of recommendations is needed to overcome concerns relating to the methodological limitations already discussed and it can also aid in helping the scientific community come to a consensus about the validity of specific gene-diet interactions. There needs to be a level of scientific rigour and standard that all studies examining interactions should aspire towards, which may be also instigated by such a list. Although several authors have made suggestions about reporting of interactions and publication of null findings,[141,142] currently a set of consistent recommendations does not exist for journals and readers to a) establish standards of practice in examining interactions and b) to help with evaluating the quality of an interaction study.

A set of guiding principles does currently exist for GWAS studies: 'Strengthening the Reporting of Genetic Association Studies (STREGA)- An Extension of the STROBE Statement.'[143] Boffetta et al. also published a proposed guideline for assessing gene-environment interactions that may help contribute to this effort.[144] Until one becomes available for gene-diet interactions, to aid my evaluation of the literature, I developed a 'quick-to-use' checklist informed by the following: Boffetta,[144] STREGA,[143] recommendations for evaluating GWAS studies for systematic reviews from HuGENet,[145] recommendations by Pearson and Manolio[146] on how to interpret a GWAS as well as concepts relevant to nutritional epidemiology (Appendix A).

| Table 0-2: Types of approaches used to study gene-diet interactions | | | | |
|---|---|---|---|---|
| **Type of method** | **What is this?** | **Example** | **Strength** | **Limitation** |
| Candidate gene-diet interaction studies | Examines one or more genetic variants (e.g. candidate genes or haplotypes), which may have often been chosen because of a biological hypothesis. This has been the predominant method of published gene-diet interaction studies to date. Alternatively, using single candidate genes identified in GWAS meta-analyses of traits of interest. | The interaction between folate status and a single C677T polymorphism in the MTHFR genetic locus on risk for stroke. This SNP reduces enzyme efficiency for processing folic acid within the folate-mediated one carbon metabolism, thereby lowering circulating folate levels and increasing homocysteine. Individuals with a T risk allele (CT or TT genotype) were found to have higher stroke risk only if residing in countries with low levels of folate in the food supply (i.e. in Asia, where there are no policies in folic acid fortification). This suggests that the effects of MTHFR on stroke risk may only become apparent under conditions of low folate intake.[147] | Simple to examine | The effect size of individual SNPs is sometimes too small to detect. Often it is difficult to determine the causal variant. |
| Genetic risk score-diet interaction studies | GRS are where several individual SNPs are combined. | The interaction between sugar-sweetened beverage (SSB) intake and a GRS composed of 32 SNPs associated with BMI, at genome-wide-significance ($p<5x10^{-8}$) on the risk for obesity. Researchers found that for individuals with a high genetic predisposition for obesity (high BMI GRS), having higher intakes of SSB led to a marked higher risk for obesity than compared to individuals with lower intakes of SSB.[148] | Combining multiple SNPs can improve the variance explained by 'genes' on predicting disease, particularly for common diseases with low penetrance SNPs. It may also be used to better reflect the polygenic nature of common multifactorial diseases such as T2D.[149] | Individual genetic variants may have differing individual interactions with dietary exposures on T2D, which may be masked when aggregated.[150] Also, this aggregation may make it more difficult to evaluate the underlying aetiology. |

| | | | | |
|---|---|---|---|---|
| Genome-environment-wide-interaction studies (GEWIS) | GEWIS studies apply a hypothesis-free approach to examining gene-diet interactions. It does this by examining the interaction between a dietary factor and every genetic variant in the genome, for an outcome of interest. Few GEWIS studies in nutritional genomics have been conducted because the extremely large sample sizes that they require have until recently been unfeasible. | A GEWIS was conducted for alcohol intake and cigarette smoking and colorectal cancer. An interaction was identified between alcohol intake and variants in the 9q22.32/HIATL1 region (p=1.76E-08 ;permuted p=3.51E-08). This suggests that the effect of alcohol intake on colorectal cancer risk may differ by genetic subgroups but results are yet to be replicated.[151] | Given the limited reproducibility of many candidate gene-diet interaction studies, it has been hoped that GEWIS may be able to help identify novel interactions at unexpected genetic loci.[35] | Methods to overcome the statistical power limitations with high multiple testing and the need for efficient methods to sieve through results have been proposed.[35] In their reviews, they suggest methods such as a joint meta-analysis with 2 degrees of freedom and others that prioritise variants through multi-stage analyses. The latter is less favourable because interaction testing should not be limited to only genetic variants identified in GWAS of marginal gene effects. |
| Dietary interventions that recruit participants by genotype (also known as 'genotype-based recall' or 'recall by genotype' | Participants undergo a dietary intervention to investigate an outcome but are recruited based on their genotype for a genetic variant of interest. Often this study design is used to clarify specific gene-diet interactions that have been widely | An example of a study examined the interaction between fat intake and APOE on blood lipids. The researchers discovered that there was a significant difference in how participants' triglyceride responded to the 3 dietary interventions, depending on their genotype.[115] | Useful to confirm the function of disease-associated genetic variants. Enables more statistical power in the context of a smaller number of participants compared to a | Conditional on the identification of robust interactions and therefore genetic variants to recruit by. Several ethical concerns prevail, with the primary |

| | | | | |
|---|---|---|---|---|
| studies) | but inconsistently reported in observational studies. | | standard RCT. | concern being the potential to violate participant's privacy and right not to know their genetic information. More is discussed in a review.[152] |
| Individual participant meta-analysis of gene-diet interactions | Given the small effects of gene-diet interactions, it is often difficult to investigate these interactions in studies with small populations. To overcome this, several research groups have come together to answer a shared research question using a meta-analysis approach where results are pooled. | CHARGE nutrition consortium, an offshoot from the Cohorts for Heart and Aging Research in Genetic Epidemiology consortia which were originally set up to investigate genetic variants associated with coronary related outcomes. An example has been their investigation into GRS and dietary pattern interactions in obesity where they did not find any interactions.[153] | By adopting a standardised analysis plan, this tries to ensure that there is as much similarity in the analysis as possible. This includes harmonisation of exposure measurements and analysis between studies. Therefore, this *de novo* interaction meta-analyses have been favoured over literature based approaches.[154] | Difficult to pool studies with heterogeneous designs, especially in the measurement of dietary exposures. |

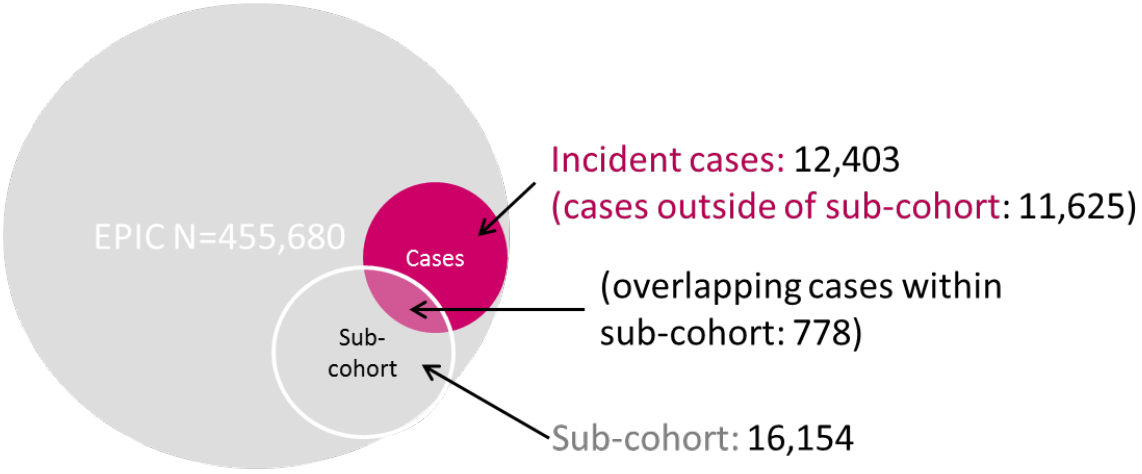
## 3.2 The EPIC-InterAct study

### *3.2.1 Cohort description*

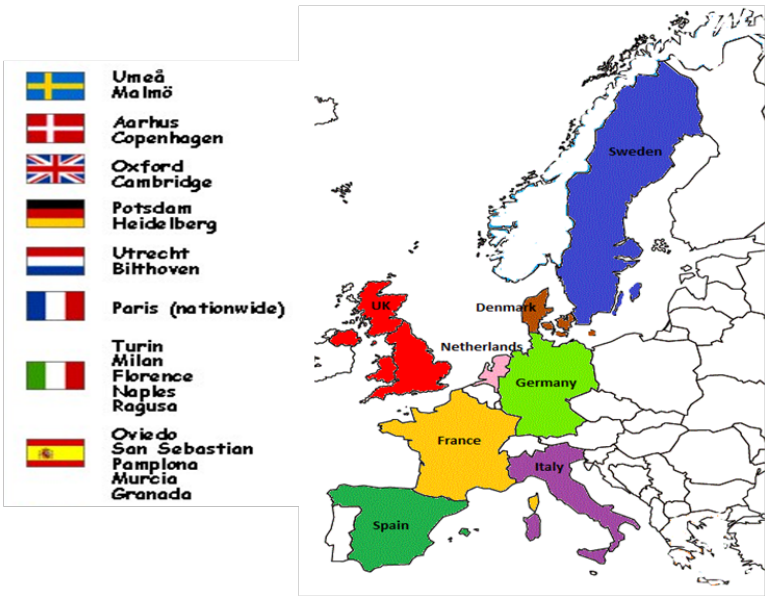
The European Prospective Investigation into Cancer and Nutrition (EPIC) is a prospective cohort study representing 519,987 participants from 23 centres in 10 European countries (Denmark, France, Germany, Greece, Italy, The Netherlands, Norway, Spain, Sweden, and the UK).[155] It was initiated in the late 1980s and is on-going. The EPIC-InterAct Study is a case-cohort study nested within EPIC, including all countries except Greece and Norway. Most participants were recruited from the general population, with some exceptions as follows: only women who were members of a health and insurance scheme for school and university employees were included in France; some blood donors were included in Turin and Ragusa (Italy) and Spain; those in Utrecht (The Netherlands) and Florence (Italy) were recruited via a breast cancer screening programme; and most participants within Oxford (the UK) were vegetarian. The design and methods have been detailed previously.[156]

Briefly, the case-cohort design of EPIC-InterAct combines the methodological advantages of i) prospective assessment of exposures which avoids recall bias and reverse causality with ii) the efficiency from needing fewer controls (Figure 3-2). From the above 8 countries eligible for InterAct (N=455,680), 340,234 participants with stored blood were eligible. A random sub-cohort of 16,835 participants was selected. After excluding individuals with prevalent T2D at baseline (1991-1997), post-censoring diabetes or unknown T2D status, a total of 27,779 EPIC-InterAct participants were included, with 12,403 incident T2D cases and a random subcohort of 16,154 participants (with the subcohort including 778 T2D cases that developed T2D during follow up, according to the design of a case-cohort study). The mean age at baseline was 55.6 years. Mean BMI in men was 29.4kg/m$^2$ and in women it was 30.1kg/m$^2$. 49.7% of participants were men.[156] Ethical approval was obtained at each centre and all participants provided informed consent.

The EPIC-InterAct study was designed to improve the understanding of the interplay between genes and the environment in the development of T2D, so it presents a richly phenotyped platform for investigating gene-diet interactions.[157]




***Total:** cases: 12,403 | non-cases: 15,376 | total: 27,779




**InterAct 8 countries |** 21 centres

**Figure 0-2: Study design of EPIC-InterAct and countries and centres represented**

### *3.2.2 Type 2 diabetes ascertainment*

Cases of clinical incident T2D were ascertained in each centre using multiple sources for verification. This included information about the participant from at least two of the following sources and was recorded at a date after the baseline visit: self-report, linkage to primary-care registers, secondary-care registers, medication use, hospital admissions and mortality data. In Denmark and Sweden, cases were considered verified because they were ascertained via local and national diabetes and pharmaceutical registers. Censoring occurred at the date of diagnosis, end of 2007 or date of death, whichever was the earlier.

### *3.2.3 Assessment of dietary intake*

Habitual self-reported dietary data was collected at baseline using country-specific self or interviewer-administered dietary questionnaires, which were developed and validated within each country to estimate usual intake of participants during the past 12 months. This contained up to 260 food items. They were either quantitative or semi-quantitative. The semi-quantitative food frequency questionnaire (FFQ) asked participants to choose from nine options for frequency of consumption from 'never or less than once per month,' to '6+ per day.' For example, the UK based FFQ in the EPIC-Norfolk study included major food groups and contained foods including meat, fish and poultry; bread and savoury biscuits; cereals (including porridge); potatoes, rice and pasta (including composite foods such as lasagne); dairy products, eggs and egg products and fats (such as butter or vegetable oil spreads); sweets and snacks (including cakes, ice-cream, chocolate, candy, chips and nuts); soups, sauces and spreads; drinks (including alcoholic and non-alcoholic beverages); fruit (including fresh, tinned and dried) and vegetables (including legumes and pulses). It concluded with open-ended questions asking about items relevant to consumption including the type of milk used, type of fat, the frequency of fried food consumption, takeaway, cooking methods (e.g. grilling or roasting meat), added salt and use of dietary supplements. A copy of this is in Appendix B. Some centres collected dietary data via diet histories. The EPIC-SOFT software was used to help harmonise country-

specific data through the collection of 24-hour dietary recalls in a representative sample of EPIC (N=32,063). This was then used to help correct for systematic between-centre over- or underestimations in dietary assessment.[155] Food composition was then derived from a standardised nutrient database called the EPIC Nutrient DataBase (ENDB), which contains 550 to 1500 foods depending on the country, standardised for unit, mode of expression, definition and chemical method of analysis.[158] Country-specific FFQs demonstrated moderate validity against a 16 day weighed food record and biomarkers for energy (doubly labelled water) and protein intake (urinary nitrogen) (Spearman correlation coefficients of 0.43 to 0.55 for macronutrients).[159–161]

#### *3.2.4 Measurement of plasma phospholipid fatty acids*

Objective dietary measures may help to overcome recall and reporting bias. The fatty acid composition of plasma phospholipids were available and could be used as nutritional biomarkers of n3, n6, total PUFA and individual and total SFA, represented as a percentage of total measured phospholipid fatty acids (mol%).[162] They were profiled using a validated high-throughput automated method, including hydrolysis and trans-methylation to produce volatile free fatty acids before separation by gas chromatography, as previously described.[163] Samples were processed randomly and laboratory staff blinded during analysis.

#### *3.2.5 Other measurements*

All participants completed baseline health and lifestyle questionnaires to obtain data on diet, physical activity, smoking, family history of diseases and socio-economic status variables and trained nurses collected data on anthropometry (e.g. weight, height, body circumferences), clinical measures (e.g. blood pressure) and biological samples (e.g. blood).[155] In France and Oxford, weight and height were self-reported by some participants.

#### *3.2.6 Genetic data*

Methods for DNA extraction from blood samples and genotyping have been previously described.[156] Genotyping was performed on the Illumina 660W-Quad

BeadChip or Illumina HumanCore Exome chip arrays and harmonised, with imputation to the Haplotype Reference Consortium using IMPUTE v2.3.2. All SNPs met quality control criteria for genotyping call rate (≥95%) or were well imputed (info≥0.99). After these evaluations, 22,492 participants had genetic data available for analysis. Samples that were genotyped on the 660 W-Quad BeadChip were randomly chosen, with the number of participants from each centre being proportional to the percentage of total cases in that centre. All centres had samples genotyped with the 660 W-Quad BeadChip array except for those from Denmark. The remaining samples were genotyped on the Illumina HumanCoreExome. There were no obvious differences in participant characteristics by array.

#### *3.2.7 Statistical analysis*

Given the case-cohort design and the over-representation of cases, a modified Cox regression (Prentice-weighted) was elected to be used as a way of analysing time to event data in case-cohort study designs. Cases within and outside the subcohort were weighted differently.[156,164] Age was used as the underlying timescale.

**Specification of exposure variables:**

- Genes were modelled additively, according to an increase in risk alleles (0, 1 or 2 risk alleles) and per standard deviation for GRS.
- Dietary exposures were modelled as continuous densities of total energy intake, to avoid losing statistical power from arbitrary categorisation. For macronutrients, this was for example, per 5% of total energy intake (TEI) from carbohydrate intake and for dietary fibre per 1g/1000kcal.
- Total energy intake was calculated by summing the energy contribution from carbohydrate, fat, protein and alcohol intake.

**Model construction for interaction analyses:**

The observational nature of this research and complexity of T2D aetiology meant that understanding the possible confounders and biases was important to help

minimise the impact of these on the findings. Therefore a directed acyclic graph (DAG)[165] of the causal path from the exposures (genetic, macronutrient intake) to the outcome, T2D, was developed to visualise this (Figure 3-3).

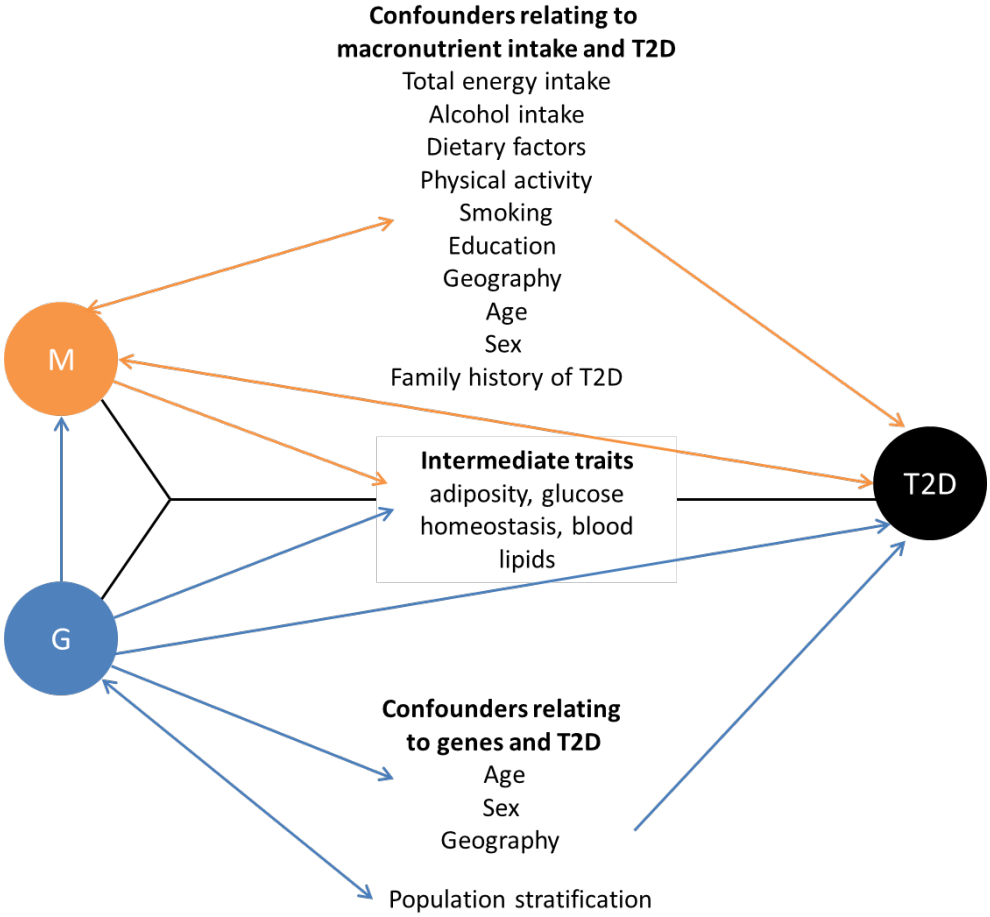



**Figure 0-3: Directed acyclic graph for the interaction between macronutrient intake and genetic variants on the risk of developing Type 2 Diabetes (T2D), including confounders for each set of exposures**. M: macronutrient exposure, G: genetic exposure.

Unless otherwise specified, the covariates were chosen if a) considered a known confounder between genetic variants and T2D, b) considered known confounders between dietary macronutrient and T2D from the literature (i.e. age, sex, geographical location, population stratification, physical activity, socioeconomic status, smoking, alcohol intake, total energy intake and BMI) or c) dietary factors that may be potential confounders (e.g. magnesium) were included within the interaction model and backward deletion was undertaken to eliminate those that did not make a notable difference to the β co-efficient of the interaction term (≤10% change) and/or induced multicollinearity until a parsimonious model was

established.[166] Regarding the rationale for the decision about the selected dietary factors to include as covariates, both 'nutrients' and 'foods or beverages,' were chosen if they had a significant association with T2D in the literature.[42] Additionally, 'food or beverage' covariates were then filtered to include only those composed minimally of macronutrients (i.e. leafy vegetables and tea and coffee). The reason for this was to preserve the variation of the macronutrient being examined. That is, if the interaction between PUFA and genetic variants on T2D was of interest, and if fish intake was included in the model (a likely confounder), it would be hard to interpret findings if the source of the variation in PUFA intake has been removed. This modelling is listed below. BMI was not adjusted in the models examining BMI GRS because the association between BMI GRS and T2D may be secondary to the association with BMI. Although a family history of T2D is a possible confounder, this data was available only in 52% of participants in EPIC-InterAct. Therefore, to avoid losing statistical power, we tested this as a sensitivity analysis, described below.

**Where appropriate, models were adjusted for the following covariates based on the DAG in Figure 3-3:**

- Age (as the underlying timescale)
- Sex
- BMI
- Centre
- Principal components for population stratification
- Physical activity (inactive, moderately inactive, moderately active and active)
- Highest level of education (none, primary school, technical/professional school, secondary school, longer education such as University)
- Smoking (never, former, current smoker)
- Sex-specific alcohol categories (none; light drinking: 0.1-6 g; moderate: women 6.1-12 g and men 6.1-24 g; heavy: women >12 g and men >24 g of alcohol/d)
- Lifestyle factors and dietary factors previously found to be associated with T2D and macronutrient intake: total energy intake, season of FFQ completion which has been associated with accuracy of recall, other macronutrients, dietary fibre, magnesium, iron, vitamin C and a selection of foods and beverages that did not contain or contained minimal amounts of macronutrients

**Plus** isocaloric macronutrient substitution

**Sensitivity analysis:**

Where a statistically significant interaction is identified, the following are a list of sensitivity analyses that were planned:

- Exclusion of those with implausible energy intakes according to thresholds used by a similar population of the Nurses' Health Study (women <500 or >3500kcal/day) and Health Professionals' Follow-up Study (men <800 or >4000kcal/day).[167]
- Adjustment for family history of T2D, only in those with available data

Separate adjustment for possible confounders and/or mediators:

- Blood lipids (total cholesterol, high-density lipoprotein, low-density lipoprotein, triglycerides) and lipid-lowering medication
- Hypertension and antihypertensive medication
- Prevalent cardiometabolic conditions (e.g. myocardial infarction or stroke)

Where additional methods were used for a particular project, these are described in the respective chapters.

Crude and multivariable-adjusted Prentice-weighted Cox regression models were constructed within country and by genotyping chip, to account for variation in genetic and dietary exposures between countries and variation in genotyping by chip. Chip-specific HRs were pooled within each country using random-effects meta-analysis. Country-specific HRs were then also combined using random-effects meta-analysis. Between-country heterogeneity was assessed with Cochran's Q statistic and quantified by the $I^2$ value and *P* for heterogeneity was derived from a Chi-squared test. *Stata* v14 (StataCorp LP, Texas, USA) was used for analysis.

**Isocaloric macronutrient substitution**

The importance of isocaloric macronutrient substitution modelling was described in section 2.5.3.2. In relevant analyses, multivariate nutrient density model with isocaloric substitution described by Willet et al., have been applied.[168] In brief, this method assumes an isocaloric model, which means that TEI is held stable. Within such a model, when energy intake is increased from a particular macronutrient (e.g. carbohydrate), this subsequently leads to a reduction in TEI from another macronutrient (e.g. fat) to ensure TEI remains constant. Therefore, the interpretation would be 'what is the effect of decreasing a proportion of energy from fat intake when replacing it with the same proportion of energy from carbohydrates intake?' Specifically, the modelling in Equation 2 helps to answer this question for T2D. The absence of 'fat density' in the model below implies a substitution effect.

$$f(T2DM) = \beta_o + \beta_C Carbohydrate\ density + \beta_P Protein\ density + \beta_A Alcohol + \sum \beta_C C_x \quad (2)$$

Please note that TEI was treated as a separate confounder adjusted within the model. Indeed, nutrient density models do capture TEI (relative measure) but they do not account for energy as a confounder which may reflect differences in body size and nutritional requirements. In fact, if TEI is a confounder, converting macronutrients into %TEI does not reduce this confounding, it simply induces confounding in the opposite direction for the macronutrient of interest because TEI will be the denominator. Moreover, as a density variable, if low TEI is associated with disease, macronutrient densities will have a positive association with the disease because of the presence of TEI in the variable even when there is no association. Therefore, to avoid this artefact, TEI was further adjusted in the model.

# *Chapter 4 Systematic literature review of gene-macronutrient interactions and T2D with attempted replication in EPIC-InterAct*

This chapter addresses objectives 1 and 2 (see Chapter 2). Here I systematically reviewed interactions reported for genetic variants and macronutrient intake in the development of Type 2 Diabetes and attempted to replicate these within EPIC-InterAct. These interactions adopted the use of candidate genetic variants.

This Chapter has been published:

**Li SX**, Imamura F, Ye Z, Schulze MB, Zheng J, Ardanaz E, et al., Interaction between genes and macronutrient intake on the risk of developing type 2 diabetes: systematic review and findings from EPIC- InterAct. *Am J Clin Nutr* 2017;106(1):263-75

## *4.1 Abstract*

**Background:** Gene-diet interactions have been reported to contribute to the development of type 2 diabetes (T2D). However, to date, few examples have been consistently replicated. To identify existing evidence for gene-macronutrient interactions on the risk of developing T2D, and to examine the reported interactions in a large-scale study.

**Methods:** We systematically reviewed studies reporting gene-macronutrient interactions and T2D that were identified using electronic databases including MEDLINE, HuGENet and the WHO clinical trials registry (to October 2015). Eligibility criteria included assessment of macronutrient quantity (e.g. total carbohydrate) or indicators of quality (e.g. dietary fibre) by use of self-report or objective biomarkers of intake. Interactions from the review were subsequently examined in the EPIC-InterAct case-cohort study (n=21,148 with 9,403 T2D cases; 8 European countries). Prentice-weighted Cox regression was used to estimate country-specific HRs, 95%CI and $p_{interaction}$, which were then pooled by random effects meta-analysis. A primary model was fitted using the same covariates as reported in the published study, and a second model adjusted for additional covariates and estimated the effects of isocaloric macronutrient substitution.

**Results:** Thirteen observational studies met eligibility criteria (n<1,700 cases). Eight unique interactions were reported to be significant between macronutrients (carbohydrate, fat, saturated fat, dietary fibre, and glycaemic load derived from self-report of dietary intake and circulating n-3 polyunsaturated fatty acids) and genetic variants in or near *TCF7L2, GIPR, CAV2* and *PEPD* ($p_{interaction}$ <0.05). We found no evidence of interaction when trying to replicate previously reported interactions. Additionally, no interactions were detected in models with additional covariates.

**Conclusions:** Eight gene-macronutrient interactions on the risk of T2D were identified from the literature. These interactions were not replicated in the EPIC-InterAct study which mirrored the analyses undertaken in the original reports. Our findings highlight the importance of independent replication of reported interactions.

### *4.2 Background*

As a first step in examining gene-diet interactions and T2D, macronutrient intake was chosen because of its importance in public health nutrition and its likelihood to interact biologically at the molecular level. A Genome-wide Complex Trait Analysis of interactions with several macronutrients across the genome also demonstrated that gene-carbohydrate interactions may explain around 25.1% (SE:14%) of the variance for fasting insulin and 24.2% (SE:13.9%) for HOMA-IR.[169] Although large confidence intervals are evident and only a few interactions reached genome-wide significance in this analysis, likely due to the small sample size of this preliminary study, the results nevertheless signify the potential value of investigating gene-macronutrient interactions in helping to understand the aetiology of T2D.

There is currently sparse confirmatory evidence for gene-macronutrient interactions and T2D. The most widely reported example is the interaction of *TCF7L2* and dietary fibre and related dietary factors (i.e. wholegrain intake), as markers of carbohydrate quality, on T2D risk.[4,112,170–172] In addition to several narrative reviews, [95,117,142] a systematic review examined lifestyle-gene interactions for T2D and highlighted the poor quality of evidence available in 2007 due to factors such as small sample size and the use of cross-sectional designs.[173] Larger prospective studies have since been published, and far more genetic loci associated with T2D risk have been identified. Furthermore, there are several important gaps in knowledge about gene-macronutrient interactions. First, past studies have not adequately controlled for confounding (e.g. by population stratification and total energy intake)[127,174,175] or considered effects of isocaloric macronutrient substitution. Second, objective biomarkers of macronutrient intake (e.g. circulating levels of polyunsaturated fatty acids-PUFA) have also not been investigated systematically. Finally, replication has been limited to date and there is potential publication bias.[142] To understand the currently available evidence and these issues, a systematic review of gene-macronutrient interactions is warranted. In synthesising summary evidence on interactions, our group had previously demonstrated that high heterogeneity

between studies prevents meaningful meta-analyses so a narrative approach was undertaken.[154]

Our first aim was to systematically review the literature relating to gene-macronutrient interactions and T2D, including both self-reported and objective markers of macronutrient intake and dietary fibre. Specifically, we planned to identify relevant interactions, assess their validity, reliability, biological plausibility, clinical relevance and to identify research gaps. Our second aim was to investigate the interactions identified from the literature-based systematic review in a large prospective study, EPIC-InterAct,[156] to address research gaps relating to replication, confounding and isocaloric macronutrient substitution.

## *4.3 Methods*

### *4.3.1 Systematic review*

This systematic review conformed to the guidelines proposed for Meta-analysis of Observational Studies in Epidemiology [176] and that of the Human Genome Epidemiology Network (HuGENet).[141,145]

Firstly, the literature was scanned for any systematic review and/or meta-analyses on this topic to assess whether a new or an update was needed for a pre-existing systematic review. Both the published literature and on-going systematic reviews on the international prospective register of systematic reviews (PROSPERO) were examined.[177]

*Selection criteria*

Studies were eligible if they reported incident or prevalent T2D as an outcome, and statistical interaction between any genetic exposure (e.g. single nucleotide polymorphisms: SNPs, genetic risk score) with macronutrient intake. Macronutrient intake included both quantity (total carbohydrate, fat and protein intake) and indicators of quality (dietary fibre, glycaemic index (GI), glycaemic load (GL), free sugars, SFA, MUFA, trans- fatty acids and PUFA, dietary

cholesterol, ratio of SFA/PUFA, linoleic acid and alpha-linolenic acid, animal and plant protein). In defining the scope for 'macronutrient intake,' the following criteria were developed:

For macronutrient quantity, it must be an essential macronutrient for physiological function and included in population dietary recommendations. This, therefore, included total carbohydrate, fat and protein. Ethanol intake was not eligible because it does not fulfil either of the above criteria.

For macronutrient quality, it can satisfy either or all of:

- included in population dietary recommendations
- evidence of an association with T2D
- is of reasonable contribution to the macronutrient quantity of interest

For example, the essential fatty acids linoleic acid (LA) and alpha-linolenic acid (ALA) contribute little to total energy intake (5.1% and 1% respectively), however, they are included for three reasons. Firstly, they contribute the most to commonly consumed dietary PUFA sources in the UK (e.g. meat, dairy, cereal products),[178] and PUFA is a quality indicator of fat intake. Secondly, often studies may represent PUFA as LA or ALA rather than total PUFA, so including these two fatty acids ensures that PUFA intake is adequately captured. Thirdly, they are essential fatty acids which have an important physiological function.

No restrictions were placed on language, age, date of publication or study design.

Exclusions included wholegrain intake. Although wholegrain intake is often associated with fibre intake because wholegrain foods are a key contributor to fibre consumption,[179] currently there is an inconsistent definition of wholegrains to accurately assess its conferred health benefits, and using it as an indicator of fibre intake is highly confounded as it contains many other bioactive compounds (B vitamins, zinc, magnesium and phytochemicals) that may affect T2D.[180,181] Moreover, it is a group of various grains such as rye, rice, oats, etc. Therefore, it cannot be categorised as a single nutrient and therefore is beyond the scope of this review. In this study, 'macronutrient' refers to both indicators of intake

(quantity and quality) as well as both methods used to assess intake (self-report and biomarker- e.g. circulating n-3 PUFA, urinary nitrogen), unless otherwise specified. Studies that assessed other forms of diabetes (e.g. type 1, gestational), examined nutrigenomics, quantitative glycaemic traits or examined the interaction between gene-lifestyle interventions without macronutrient assessment were excluded (Table 4-1).

| **Table 3-1: Systematic review selection criteria** | |
|---|---|
| **Inclusion** | **Exclusion** |
| ***Macronutrient exposure:*** | |
| **Carbohydrate quantity (grams or percentage of energy intake: %E) or quality: free sugars, non-starch polysaccharides (dietary fibre), GI, GL** | Macronutrients which will provide both caloric and nutritive value so ethanol will not be included in this review |
| **Protein quantity (grams or %E intake) or quality: meat versus plant sources** | Other environment- gene interactions e.g. physical activity, smoking, pollution; non-macronutrient: foods and dietary patterns i.e. wholegrain intake; or lifestyle interventions without macronutrient on its own |
| **Fat quantity (grams or %E intake) or quality: SFA, MUFA, trans and PUFA, dietary cholesterol, ratio of SFA/PUFA**<br>**Essential fatty acids: Linoleic acid (LA), alpha-linolenic acid (ALA), n-6/n-3 ratios** | |
| **(both self-reported and objective biomarkers of intake was eligible)** | |
| ***Genetic exposure:*** **any single variant, haplotype or combined risk score** | |
| ***Outcome:*** **incident or prevalent T2D** | Other forms of diabetes: such as type 1 diabetes, gestational diabetes, MODY |
| ***Study design*** | Nutrigenomic studies, ecological, reviews or commentaries |

*Search strategy and data extraction*

Following a pre-defined protocol, electronic searches were performed using MEDLINE, EMBASE, Cochrane Library and HuGENet up to 31st October 2015 (an example is available in Appendix C). To minimise publication bias, the following were also searched: the WHO clinical trials registry, the grey literature (e.g. GreyNet), names of key authors and diabetes trials, and hand-searched relevant reviews. Medical Subject Headings and specific terms (i.e. title, abstract and keywords) were also used wherever possible to ensure sensitivity within respective databases. Authors of three published [172,182,183] and an unpublished

study (ClinicalTrials.gov ID: NCT01168297) were contacted to either assess eligibility or collect further data to conduct the review. Subsequently two studies (n= 1 published.[183] 1 unpublished) were determined ineligible. Studies were screened by title and abstract for eligibility for full-text review. From each publication meeting eligibility criteria, the following information were extracted using an agreed data extraction form on cohort characteristics (e.g. study design, sample size, ethnicity, etc), covariates, statistical analyses, estimates of associations between 1) macronutrient intake and T2D, 2) genetic variant and T2D and 3) gene-macronutrient interactions and T2D.

*Assessment of study quality*

Assessment for confounding, bias (selection, measurement, attrition, outcome and reporting) and genetic-specific issues (genotyping quality, population stratification, multiple testing) were undertaken using a modified version of the Cochrane guidelines for non-randomised studies of interventions to incorporate genetic issues highlighted by the HuGE Network.[145,184] This broadly classified studies as being of low, moderate, serious or critical risk of bias.

Two authors (Sherly Li and Zheng Ye) independently undertook every stage of screening, selection, data extraction and quality assessment in duplicate and resolved any disagreements by a discussion with two other authors (Nita Forouhi and Robert Scott).

*Data analysis*

A narrative synthesis was undertaken. Studies were heterogeneous in reporting of interaction results, therefore for consistency, results were analysed and presented according to categories of macronutrient intake where possible. The large extent of heterogeneity made any form of standardisation impossible. For the same reason, publication bias could not be quantitatively assessed.

### *4.3.2 EPIC-InterAct Study*

To investigate the reproducibility of the statistically significant interactions identified from this systematic review, these were examined in a large-scale study (EPIC-InterAct), which had informed consent and approved ethics.[156] The cohort, genotyping and imputation, dietary assessment and estimation of macronutrient intake have already been described in the Methods Chapter (Chapter3). In addition to the general method taken for statistical analyses, described in the Methods chapter, below are some specific methods for this piece of work.

*Additional method for genetic data*

For quality control, the Hardy-Weinberg equilibrium (HWE) was estimated for the SNPs of interest within EPIC-InterAct. Theoretically, allele frequencies should be stable within a population due to the random assortment in mating if there are no perturbations. This is demonstrated in equation 3 below which shows how in a population, the sum of the homozygotes for the p allele, homozygotes for the q allele and the heterozygotes should equate to 1. Therefore, deviation from equilibrium within the healthy control group is commonly tested within genetic epidemiology and can signal possible selection bias, population stratification or genotyping error.[185] Typically, this is not tested in the cases because of obvious non-random selection based on their disease state. However, reliability has been debated because it is underpinned by the assumption that there is no perturbation in this equilibrium including genetic drift, migration, mutations, non-random mating and natural selection, which does happen in reality. There is also concern about misleading results when power for HWE testing is limited.[185] With these limitations in mind, the HWE was therefore chosen as a preliminary test in flagging possible errors.

$$p^2+2pq + q^2=1 \qquad (3)$$

*Additional methods for statistical analysis*

Additive models for all genetic variants were assumed unless previously published studies demonstrated a more appropriate alternative. Each macronutrient was categorised based on the distribution of the macronutrient intake within the subcohort sample, excluding outliers (±3 standard deviations from the mean). To account for between-country variations in dietary intake, categorisation was performed per country and then country-specific Cox regression was conducted. Since categorisation was performed in each country, the pooled category-specific ranges may appear to overlap. However, individuals are mutually exclusive within each category by country. The pooled macronutrient percentiles are given below and shown as median (minimum, maximum):

| **Macronutrient** | **Category 1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| Dietary fibre (grams/1000kcal) | 7.5 (2.4-9.5) | 9.5 (7.3-11.0) | 10.9 (8.7-12.3) | 12.4 (10.0-14.0) | 15.0 (11.5-37.0) |
| Glycaemic load (grams of carbohydrate) | 87.8 (14.1-118.5) | 123.6 (100.6-161.1) | 165.6 (132.8-436.6) | | |
| Carbohydrate (%total energy intake) | 37.2 (12.6-43.2) | 44.0 (38.4-49.1) | 50.6 (44.1-73.9) | | |
| Fat (%total energy intake) | 29.3 (5.1-34.0) | 34.7 (31.0-38.6) | 40.2 (35.7-64.9) | | |
| Circulating n-3 PUFA (%total phospholipid fatty acid) | 5.4 (0-7.5) | 7.8 (5.2-23.1) | | | |

Two approaches to modelling were taken: a 'replication model' and a 'modified model.' The 'replication model,' tried to mirror the interaction model as reported by the original publication as much as possible, to ensure comparable results. This included whether macronutrients were treated as continuous or categorical variables, whether an additive or codominant genetic model was applied, as well as selection of the effect allele and of covariates. P value for interaction was estimated with SNP and macronutrient variables treated as either categorical or continuous as per that reported in the previous publication. Whereas, the

'modified model' accounted for isocaloric macronutrient substitution and additional confounders that may bias the interaction results. This was to address the previously demonstrated risk of inflation of type I error under the inadequate control of environmental or genetic confounders.[127] Given that physical activity, education, smoking and alcohol intake are known confounders between macronutrient intake and T2D, these were included in the model. Similarly, population stratification is a well-recognised confounder for genetic exposures, as demonstrated by Sul et al., and should be included in gene-macronutrient interaction analyses.[126] In addition to confounding, the issue of isocaloric macronutrient substitution has already been discussed and have been thus far neglected in the observational literature on gene-macronutrient interactions and cardiometabolic diseases. Therefore, to improve the precision of our estimates and aid interpretation of results on macronutrient intake, modelling was based on that described in the Methods Chapter. However, please see Appendix D for a full covariate list for each respective macronutrient. P value for interaction was estimated with SNP and macronutrient variables both treated as continuous variables unless otherwise specified.

For two replication analyses, we excluded EPIC-InterAct centres Potsdam and Malmo, which contributed to previous analyses.[4,186] This included the interaction between *CAV2* and total fat and SFA which was identified by Fisher et al., and conducted in EPIC-Potsdam, so Potsdam was excluded from the present EPIC-InterAct analysis. Also, the interaction between dietary fibre and *TCF7L2* was identified by Hindy et al., in a population including participants from Malmo, so this centre was excluded in the present EPIC-InterAct analysis.

For sensitivity analysis, the interaction between dietary fibre and *TCF7L2* was examined using a dietary fibre variable which was calibrated to account for different degrees of measurement bias between centres. This is because some centres asked about non-white bread rather than specifically wholegrain bread intake. This may have resulted in different degrees of misclassification of dietary fibre, by centre, if non-white bread included for example malted brown bread which is not higher in fibre than white bread. Certainly, previous EPIC authors have shown that measurement bias in protein intake differs in magnitude and

direction across countries.[187] To try and address this differential measurement bias, analysis using a calibrated dietary fibre variable was examined. This was created by calibrating dietary fibre estimated from diets measured by FFQ with dietary fibre estimated from a single measure of 24-hour recall.[188]

The software QUANTO (http://biostats.usc.edu/Quanto.html) was used to evaluate the statistical power for EPIC-InterAct to detect the specific interactions of interest.

Complete case analyses were undertaken so that those with missing macronutrient intake, genetic data or covariates were excluded. Stata version 14 (StataCorp LP, Texas, USA) was used for all analyses with p-value for interaction of <0.05 judged as statistically significant on the basis that each interaction was considered an independent replication attempt.

## *4.4 Results*

### *4.4.1 Systematic review*

4003 publications were screened and 13 publications were included in this review (Figure 4-1). Participants had a mean age of 50 years and were on average overweight (mean BMI: 27kg/m$^2$). Characteristics are provided in Table 4-2. Study designs included four which were cross-sectional,[182,189–191] two case-control,[172,192] one family-based association study [193] and six prospective (cohort or case-cohort) studies.[4,170,186,194–196] Sample sizes ranged from 805 [191] to 24,840 [194] participants (n=165 to 1,649 cases). Dietary assessment method included self-reported diet (n=12) and one which measured erythrocyte phospholipid n-3 PUFA.[192] Across the studies examined, all macronutrients were represented except for protein quality (animal or plant protein). Interactions were examined with SNPs from nine candidate genetic loci (*TCF7L2, GIPR, IRS1, PPARγ, APOA2, CAV2, FABP1/2/3/4, PGC-1α, PEPD*) and a GRS comprising variants in 15 T2D-associated loci.[189] There was high heterogeneity in macronutrient categorisation, which foods used to estimate certain macronutrients such as cereal fibre, genetic model, statistical interaction method and reporting was evident.

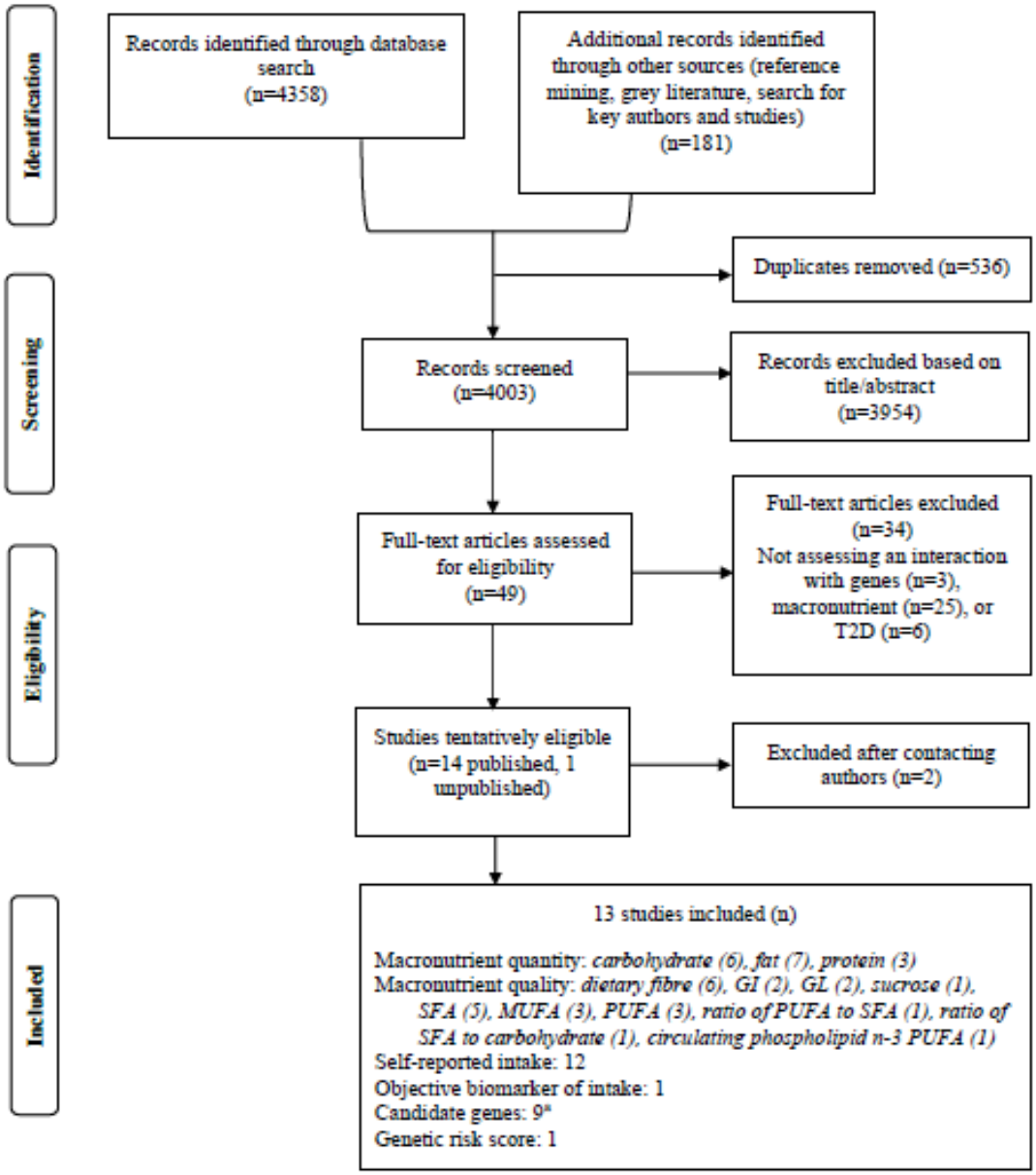



**Figure 3-1: Flow diagram of the systematic review for gene-macronutrient interactions and the risk of Type 2 Diabetes**

Abbreviations: GI: glycaemic index; GL: glycaemic load; SFA: saturated fatty acid; MUFA: monounsaturated fatty acid; PUFA: polyunsaturated fatty acid. Numbers are not mutually exclusive.

[a] not including exploratory studies which examined many candidate genetic variants

### *Gene-macronutrient interactions from the systematic review*

Eight interactions between SNPs and macronutrients were reported to be significant. These included interactions between two SNPs in the Transcription Factor 7-Like 2 (*TCF7L2)* gene with dietary fibre,[4,170] another *TCF7L2* variant

with GL,[172] a genetic variant in Gastric Inhibitory Polypeptide Receptor (*GIPR)* with total fat and carbohydrate intake,[194] a genetic variant in Caveolin 2 *(CAV2)* with total fat and SFA,[186] and a genetic variant in Peptidase *(PEPD)* with erythrocyte phospholipid n-3 PUFA.[192] These are summarised in Table 4-2 (magnitude of effects in Appendix E) and are described next.

Several studies examined variants in or near *TCF7L2*, the common variant with the strongest association with T2D.[27] This gene encodes for a high mobility group box-containing transcription factor with a role in the Wnt signalling pathway and blood glucose homeostasis. In this current review, the interaction with dietary fibre was the most widely examined (by four studies), albeit inconsistently replicated. One study reported that the effect of the T allele of rs7903146 (within *TCF7L2*) on T2D risk was significantly increased with higher intakes of total dietary fibre,[4] which was corroborated by another study investigating cereal fibre.[170] However, the results of two other studies were discordant.[172,189] Moreover, Cornelis et al., observed another interaction among US women, where T allele carriers demonstrated increased odds of T2D with diets higher in GL.[172] Dietary fibre was also the subject of interaction analyses with the following genes but no statistical interaction was found, *GIPR*, *IRS1* or *PPARγ* or a T2D GRS.

One study reported that carriers of the A allele for SNP rs10423928 (*GIPR)* had a lower 12-year incidence of T2D only if also consuming a diet higher in fat or lower in carbohydrate.[194] *GIPR* was chosen based on the hypothesis that it encodes the receptor for the incretin hormone gastric inhibitory polypeptide,[194] which stimulates insulin release in the presence of elevated glucose.

Another study followed up *CAV2* (rs2270188) for interaction with fat intake after exploratory analysis. This gene encodes a protein found on the surface of caveolae (small invaginations of cellular plasma membranes) and may be involved in lipid metabolism. It has not been previously associated with T2D (OR: 0.99, 95%CI 0.97, 1.01, p= 0.49).[5] When individuals with the GG genotype were compared to those with TT genotype, those with the TT genotype had a higher risk of T2D when consuming diets higher in total fat and SFA.[186]

In a Chinese case-control study, Zheng et al. reported an interaction between circulating n-3 PUFA and *PEPD,*[192] which encodes a peptidase involved in proline recycling and collagen production. Within the gene, *PEPD,* rs3786897 has been associated with T2D in Asians.[197] It has not been associated with T2D in Europeans within DIAGRAM (OR of T2D: 0.97; 95%CI: 0.95,0.99; p:0.04). Individuals with a GA or AA genotype were found to be at higher risk of T2D, compared to those with a GG genotype, only among adults possessing lower levels of n-3 PUFA (≤ 5.33% of total circulating phospholipid fatty acids).

Three large cohort studies examined protein intake and SNPs in/near *TCF7L2, GIPR and IRS1* for interaction on T2D. However, no interactions were reported for protein intake.

Four studies investigated the interaction between fatty acid intake and SNPs in or near *PPARγ* on T2D risk but none were statistically significant.

*Assessment of risk of bias and quality of evidence*

All studies included in the review were observational and rated either at moderate (n=8) or serious risk of bias (n=5) (Appendix E for more information). Of the 6 studies reporting interactions, three did not account for multiple testing corrections ($\alpha<0.05$) when examining several macronutrients and/or SNPs (e.g. an exploratory study examining 64 SNPs with 4 fatty acids; total 256 tests).[4,186,194] Two studies which published a statistically significant interaction included accompanying replication results.[186,190] Many studies do not adjust for known confounders. Confounders including total energy intake, physical activity and population stratification were frequently ignored,[192,196,198] in particular the last was considered in only one study.[182] Other concerns included the validity and reliability of the dietary measurement tool [170,189,191,196] and possible selective analysis and reporting.[182,191]

| Table 3-2: Summary of all eligible studies within the systematic review of gene-macronutrient interactions and T2D | | | | | |
|---|---|---|---|---|---|
| **Genetic locus and variant** | **Reference and study name (n=cases/total)** | **Country and population characteristics (age in years, BMI in kg/m²)** | **Macronutrient** | **Interaction results (p value for interaction)** | **Quality assessment** |
| ***TCF7L2***<br>rs7903146 | Hindy, 2012 [4]<br>(MDCS cohort<br>n=1649/24799) | Sweden, European<br>Age: 58.1±7.6<br>39%M<br>BMI: 25.7±3.9 | Carbohydrate<br>Fat<br>Protein<br>Total fibre | Dietary fibre * *TCF7L2*<br>(p=0.049).<br>↑ fibre ↓ T2D risk in CC genotype.<br>↑ fibre ↑ T2D risk in T allele carriers. | Moderate<br>Large sample size with validated measurement tools and adjustment for confounders. Multiple testing not accounted for (4 tests). |
| rs7903146<br>rs4506565 | Wirstrom, 2013 [170] (SDPP cohort<br>n=165/5477) | Sweden, European<br>Age: 47.2<br>42%M | Cereal fibre | rs7903146 (p=0.005)<br>↑ fibre ↓ T2D risk in CC genotype.<br>rs4506565 (p=0.006)<br>↑ fibre ↓ T2D risk in AA genotype. | Moderate<br>Exposure is not comprehensively measured and study with small sample size. |
| rs12255372 | Cornelis, 2009 [172]<br>(NHS case-control<br>n=1140/3055) | USA, European<br>Age: 47.5±6.9<br>0%M<br>BMI: 24.5±4.6 | Carbohydrate<br>GI<br>GL<br>Cereal fibre | GL * *TCF7L2* (p=0.003)<br>*interaction disappeared after adjusting for family history (p=0.13) | Moderate<br>Well conducted study. |
| ***GIPR***<br><br>rs10423928 | Sonestedt, 2012 [194] (MDCS cohort<br>n=1541/24840) | Sweden, European<br>Age: 58±7.7<br>39%M<br>BMI: 25.7±4 | Carbohydrate<br>Fat<br>Protein<br>Fibre<br>Sucrose | Carbohydrate * *GIPR* (p=0.001)<br>Fat * *GIPR* (p=0.002)<br>↑ fat and ↓ carbohydrate reduce T2D in A allele carriers. The opposite composition for T allele carriers. | Moderate<br>Large sample size, validated measurement tools and comprehensive analyses, with adjustment for key confounders. Multiple testing not accounted for (5 tests). |
| ***IRS1***<br>rs2943641 | Ericson, 2013 [195]<br>(MDCS cohort<br>n=1567/24841) | Sweden, European<br>Age: 58±7.7<br>39%M<br>BMI: 25.6±4 | Carbohydrate<br>Fat<br>Protein<br>Fibre | No significant interaction<br>(p=0.59)<br>(p=0.40)<br>(p=0.28)<br>(p=0.92) | Moderate<br>Large sample size with comprehensive dietary measurement. Multiple testing was not accounted for. |

| Genetic locus and variant | Reference and study name (n=cases/total) | Population characteristics (age in years, BMI in kg/m$^2$) | Macronutrient | Interaction results (p value for interaction) | Quality assessment |
|---|---|---|---|---|---|
| rs7578326<br>rs2943641 | Zheng, 2013 [190]<br>(GOLDN, BPRHS cross-sectional n=419/1664) | USA, Hispanic, African, European, Native American<br>Age: 53.5±13.2<br>38%M<br>BMI: 30.3±6.4 | Carbohydrate<br>Fat<br>SFA<br>MUFA<br>SFA: Carbohydrate ratio<br>GI GL | No significant interaction (p= not available) | Moderate<br>Conducted replication and meta-analysis. Well adjusted for confounding. Small sample size. |
| ***PPARγ***<br>Pro12Ala/rs1801282 1431C>T | Lamri, 2012 [196]<br>(DESIR cohort n=191/4676) | France, European<br>Age: 46.8±10<br>49%M<br>BMI: 24.7±3.8 | Fat | No significant interaction (p=0.05) | Serious<br>Dietary questionnaire with limited food items may cause measurement bias and misclassification of fat intake. Residual confounding is likely due to lack of adjustments for key confounders (total energy intake, physical activity, etc). |
| Pro12Ala | Cornelis, 2009 [172]<br>(NHS case-control n=1140/3055) | USA, European<br>Age: 47.5±6.9<br>0%M<br>BMI: 24.5±4.6 | Carbohydrate<br>GI<br>GL<br>Cereal fibre | No significant interaction (p= not available) | |
| Pro12Ala | Nelson, 2007 (37)<br>(GENI study: family-based association analysis n=736/1318) | USA, Hispanic, European<br>Age: 40.9±19.4<br>43%M<br>BMI: 30.5±6.6 | PUFA<br>SFA<br>MUFA<br>PUFA: SFA ratio | No significant interaction (p= not available) | Serious<br>Family-based association test conducted. Several key confounders (energy intake, BMI, physical activity, etc) not considered and possible reporting bias. |

| Genetic locus and variant | Reference and study name (n=cases/total) | Population characteristics (age in years, BMI in kg/m$^2$) | Macronutrient | Interaction results (p value for interaction) | Quality assessment |
|---|---|---|---|---|---|
| Pro12Ala<br>63 SNPs examined with only CAV2 and PPARγ taken forward for confirmation | Fisher, 2011 [186]<br>(EPIC-Potsdam case-control 576, case-cohort 806/2864) | Germany, European<br>Age: 50.4±8.9<br>42%M<br>BMI: 26.7±4.6 | Fat<br>SFA<br>MUFA<br>PUFA | No significant interaction<br>(p=0.32)<br>(p=0.08)<br>(p=0.29)<br>(p=0.07) | Moderate<br>Robust assessment, using a novel interaction analysis approach that maximises power. Multiple testing was not accounted for in exploratory analysis (256 tests). |
| ***APOA2***<br><br>−265T>C | Corella, 2011 [182]<br>(PREDIMED and SNHS are both cross-sectional n=825/2830) | Singapore, Asian<br>Spain, European<br>Age: 44.1±16<br>41%M<br>BMI: 25±5.3 | SFA | No interaction reported (p= not available) | Moderate<br>Potential reporting bias evident. |
| ***CAV2***<br>rs2270188<br>63 SNPs | Fisher, 2011 [186]<br>(EPIC-Potsdam case-control 576, case-cohort 2864) | Germany, European<br>Age: 50.4±8.9<br>42%M<br>BMI: 26.7±4.6 | Fat<br>SFA<br>MUFA<br>PUFA | Fat * *CAV2* (p=0.02)<br>SFA * *CAV2* (p=0.002)<br>↑ fat and SFA ↑ T2D among TT genotype.<br>(confirmatory analyses) | |
| ***FABP1/2/3/4***<br><br><br><br>rs2197076<br>12 SNPs | Mansego, 2012 [191] (Hortega, Segovia replication cross-sectional n=174/2022) | Spain, European<br>Age: 52.8±11.2<br>45%M<br>BMI: 27.5±4.1 | Fat<br>SFA<br>PUFA | No significant interaction post multiple testing corrections<br>(p=0.03) | Serious<br>Replication conducted. The interaction was neither examined in 1/3 of the population (without reason) nor examined in SNPs without the main effect on T2D. Most confounders were not accounted for and selective reporting was evident. |

| Genetic locus and variant | Reference and study name (n=cases/total) | Population characteristics (age in years, BMI in kg/m$^2$) | Macronutrient | Interaction results (p value for interaction) | Quality assessment |
|---|---|---|---|---|---|
| ***PGC-1α***<br><br>Gly482Ser<br>Thr612Met<br>Thr528Thr | Nelson, 2007 (37) (GENI study: family-based association analysis n=736/1318) | USA, Hispanic, European<br>Age: 40.9±19.4<br>43%M<br>BMI: 30.5±6.6 | PUFA<br>SFA<br>MUFA<br>PUFA: SFA ratio | No significant interaction (p= not available) | |
| ***PEPD***<br><br>rs3786897<br>9 SNPs examined | Zheng, 2015 [192] (case-control, n=622/915) | China, Asian<br>Age: 51.1±13.2<br>51%M<br>BMI: 24.5±2.7 | Circulating erythrocyte membrane phospholipid n-3 PUFA | n-3 PUFA * *PEPD* (p=0.027)<br>↓ n-3 PUFA ↑ T2D among A allele carriers<br>↑ n-3 PUFA is not associated with T2D among A allele carriers | Serious<br>Several likely confounders may explain the associations observed that were not adjusted for (BMI, blood lipid status, etc) and incomplete reporting of participant characteristics. |
| **GRS**<br>Based on 15 T2D genetic loci, weighted score | Villegas, 2014 (32) (NHANES cross-sectional n=1337/13120) | USA, European<br>Age: 51.1±13.2<br>51%M<br>BMI: 24.5±2.7 | Carbohydrate<br><br>Fibre | No significant interaction (Non-Hispanic whites p=0.53)<br>(p=0.09) | Serious<br>Dietary measurement tool was not validated and how T2D status was obtained was not described, therefore bias may be likely. The study reported a lack of statistical power to detect interactions. |

For the magnitude of effects (e.g. OR and 95%CI), refer to Appendix E.

**Abbreviations:** M: male, GL: glycaemic load, GI: glycaemic index, MDCS: Malmo Diet and Cancer Study, SDPP: Stockholm Diabetes Prevention Program, NHS: Nurse's Health Study, GOLDN: Genetics of Lipid Lowering Drugs and Diet Network, BPRHS: Boston Puerto Rican Health Study, GENI: Gene-Environment Interactions study, EPIC: European Prospective Investigation into Cancer and Nutrition-Potsdam, PREDIMED: Prevención con Dieta Mediterránea trial, SNHS: , Singapore National Health Survey, NHANES: National Health and Nutrition Examination Survey.

### *4.4.2 Findings in EPIC-InterAct*

The population of EPIC-InterAct used for this analysis was broadly similar to the average of the cohorts from the systematic review. The mean age at baseline was 52.3years and 55.7years in non-cases and cases of T2D, respectively and overweight (mean BMI: 25.8kg/m² and 29.7kg/m² in non-cases and cases, respectively) (Table 4-3). Associations between the SNPs we attempted to replicate and T2D were comparable with the genome-wide meta-analysis of genetic variants for T2D previously published (Table 4-4).[5]

**Table 3-3: Baseline characteristics of the participants from EPIC-InterAct (based on complete case analysis)**

| Characteristics | Mean/% | SD/n | Mean/% | SD/n | % missing |
|---|---|---|---|---|---|
| | **Subcohort non-cases** | | **Incident cases** | | |
| **Number** | 11745 | | 9403 | | Total: 21148^ |
| **Average years of follow-up** | 12.2 | | 6.9 | | |
| **Age (y)** | 52.3 | 9.2 | 55.7 | 7.6 | 0 |
| **Sex (%male)** | 37.9 | 4452 | 50.0 | 4704 | 0 |
| **PA level (%)** | | | | | 1.1% |
| inactive | 22.4 | 2636 | 29.5 | 2778 | |
| moderately inactive | 33.7 | 3956 | 33.0 | 3102 | |
| moderately active | 23.1 | 2712 | 20.3 | 1909 | |
| active | 20.8 | 2441 | 17.2 | 1614 | |
| **Highest school level (%)** | | | | | 1.7% |
| none | 6.8 | 802 | 9.2 | 862 | |
| primary school | 32.1 | 3772 | 41.8 | 3928 | |
| technical/professional | 24.4 | 2863 | 24.9 | 2339 | |
| secondary school | 15.1 | 1773 | 10.7 | 1005 | |
| longer education (inc. university) | 21.6 | 2535 | 13.5 | 1269 | |
| **Family history of diabetes (%)** | | | | | 47.8% |
| Yes | 18.1 | 1105.0 | 35.9 | 1748.0 | |
| No | 81.9 | 5006.0 | 64.2 | 3128.0 | |
| **BMI (kg/m²)** | 25.8 | 4.0 | 29.7 | 4.7 | 0.6% |
| **HbA1c (mmol/mol)** | 36.0 | 4.2 | 43.6 | 10.6 | 2.0% |
| **Total cholesterol (mmol/L)** | 5.9 | 1.1 | 6.2 | 1.2 | 5.2% |
| **LDL (mmol/L)** | 3.8 | 1.0 | 4.0 | 1.0 | 7.7% |
| **HDL (mmol/L)** | 1.5 | 0.4 | 1.2 | 0.4 | 5.2% |
| **TAGs (mmol/L)** | 1.3 | 0.9 | 2.0 | 1.3 | 5.2% |
| **Smoking status (%)** | | | | | 1% |
| never | 46.6 | 5477 | 40.4 | 3798 | |

| | | | | | |
|---|---|---|---|---|---|
| former | 27.1 | 3178 | 31.2 | 2935 | |
| current smoker | 26.3 | 3090 | 28.4 | 2670 | |
| **Energy intake (kcal/d)** | 2137.3 | 635.1 | 2176.0 | 674.9 | 3.4%* |
| Fat (%E) | 34.8 | 5.8 | 34.8 | 6.0 | 3.4%* |
| Saturated fat (%E) | 13.4 | 3.4 | 13.3 | 3.5 | 3.4%* |
| Monounsaturated fat (%E) | 13.1 | 3.4 | 13.1 | 3.3 | 3.4%* |
| Polyunsaturated fat (%E) | 5.5 | 1.8 | 5.6 | 1.9 | 3.4%* |
| Protein (%E) | 16.9 | 3.0 | 17.2 | 3.2 | 3.4%* |
| Animal protein (%E) | 10.5 | 3.3 | 11.0 | 3.4 | 3.4%* |
| Plant protein (%E) | 5.0 | 1.3 | 4.9 | 1.3 | 3.4%* |
| Carbohydrate (%E) | 44.0 | 7.0 | 43.6 | 7.2 | 3.4%* |
| Fibre (g/1000kcal) | 10.9 | 3.1 | 10.7 | 3.2 | 3.4%* |
| Fibre- vegetable (g/1000kcal) | 2.0 | 1.3 | 1.9 | 1.4 | 0.0% |
| Fibre- fruit (g/1000kcal) | 2.1 | 1.7 | 2.1 | 1.7 | 1.3% |
| Fibre- cereal (g/1000kcal) | 4.2 | 2.1 | 4.2 | 2.1 | 0.0% |
| Glycaemic Load (g of carbohydrate) | 131.6 | 45.4 | 132.8 | 47.1 | |
| **Alcohol intake (sex-specific g/d)** | | | | | 3.4%* |
| none | 13.6 | 1601 | 16.4 | 1539 | |
| light drinking (0.1-6g) | 34.9 | 4093 | 33.2 | 3123 | |
| moderate drinking (men: 6.1-24g, women: 6-12g) | 23.6 | 2770 | 24.3 | 2286 | |
| heavy drinking (men>24g, women>12g) | 27.9 | 3281 | 26.1 | 2455 | |
| **Circulating fatty acid (% total phospholipid fatty acid)** | **Subcohort non-cases** | | **Total incident cases** | | 1.3% Total: 22273^ |
| | 12336 | | 9937 | | |
| total PUFA | 42.5 | 3.5 | 41.9 | 3.5 | |
| n3 PUFA | 6.7 | 2.0 | 6.8 | 2.1 | |
| n6 PUFA | 35.7 | 3.7 | 35.1 | 3.7 | |

* represents % missing based on total sample of 27,779

All other missing % based on a sample of 21900 (after exclusion of individuals with missing macronutrient n=736 and genetics data n=5,287). ^ total sample size and descriptive statistics: based on complete case analyses after exclusion of those with missing exposure and covariates. Abbreviations: PA: physical activity; BMI: body mass index; LDL: low-density lipoprotein; HDL: high-density lipoprotein; TAGs: triglyceride; g: grams; E: total energy intake; SFA: saturated fatty acids, MUFA: monounsaturated fatty acids; PUFA: polyunsaturated fatty acids

There were no significant interactions for any of the replication analyses in EPIC-InterAct that were comparable to the model specifications in the published literature. Figure 4-2A shows that compared to the original report ($p_{interaction}$= 0.049) (1,649 cases of T2D/24,799 total),[4] in EPIC-InterAct we failed to replicate the significant interaction between *TCF7L2* rs7903146 and dietary fibre intakes on incident T2D ($p_{interaction}$ = 0.97) (8,012 cases of incident T2D/ 18,292

total). The covariates included in each model are detailed in the figure legend. No interactions were observed in EPIC-InterAct by sub-types of dietary fibre (cereal, vegetable or fruit fibre) ($p_{interaction}$ ≥0.27) (Figure 4-3). Figure 4-2B shows no replication of the interaction between *TCF7L2* and GL on T2D ($p_{interaction}$ = 0.58) as previously detected by Cornelis et al.,[172] Similarly, in EPIC-InterAct no significant interactions were detected between rs10423928 (in *GIPR*) and carbohydrate or fat intake on incident T2D ($p_{interaction}$ = 0.79 and 0.25, respectively)[194] (Figure 4-4). At *CAV2*, where an interaction had been reported between both total fat and SFA intake with rs2270188, [186] there was no evidence to support this in EPIC-InterAct ($p_{interaction}$ = 0.76 and 0.95, respectively) (Figure 4-5). In additional analysis, however, a significant interaction was detected when Potsdam was analysed independently, the centre originally analysed in the previous publication,[186] ($p_{interaction}$ = 1.01E-6 and 0.001 for total and SFA, respectively) (Figure 4-6). The interaction between rs3786897 (within *PEPD*) and circulating n-3 PUFA reported by Zheng et al.,[192] was also not observed in EPIC-InterAct ($p_{interaction}$ = 0.58) (Figure 4-7).

**Table 3-4: EPIC-InterAct results for the association between SNPs of interest and incident T2D**

| SNP rsid | chr | nearest gene | call rate | info score | HWE | effect allele | HR for T2D | lower CI | Upper CI | p-value | $I^2$ (%) | p_het | DIAGRAM results (OR;95%CI; p) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs7903146 | 10 | *TCF7L2* | >95% | NA | 0.44 | T | 1.33 | 1.24 | 1.43 | 1.74E-16 | 53 | 0.04 | 1.39;1.35,1.42;1.2E-139 |
| rs122255372 | 10 | *TCF7L2* | >95% | NA | 0.78 | T | 1.29 | 1.20 | 1.38 | 2.37E-12 | 56 | 0.03 | 1.33; 1.30,1.37; 1.2E-12 |
| rs10423928 | 19 | *GIPR* | >95% | 1.00 | 0.4 | A | 1.03 | 0.98 | 1.08 | 0.19 | 0 | 0.49 | 1.05; 1.02,1.08; 0.00 |
| rs3786897 | 19 | *PEPD* | >95% | 1.00 | 0.03* | A | 1.06 | 1.01 | 1.10 | 0.01 | 0 | 0.64 | 0.97; 0.95,0.99; 0.04 |
| rs2270188 | 7 | *CAV2* | >95% | 0.99 | 0.48 | T | 0.96 | 0.92 | 1.01 | 0.13 | 29 | 0.20 | 0.99; 0.97,1.01; 0.49 |

SNPs imputed using Haplotype Reference Consortium

Abbreviations: info: information content metric; HWE: Hardy-Weinberg equilibrium; HR: hazard ratio; T2D: type 2 diabetes; CI: confidence interval; p_het: p-value for heterogeneity
Pooled HR for T2D, with adjustment age (=underlying time scale), sex, centre and eigenvectors (first 5 principle components for population stratification)
DIAGRAM results from Morris et al., Nat Genet, 2012
*HWE p-value threshold for rejecting the null p<0.01 (5 SNPs)

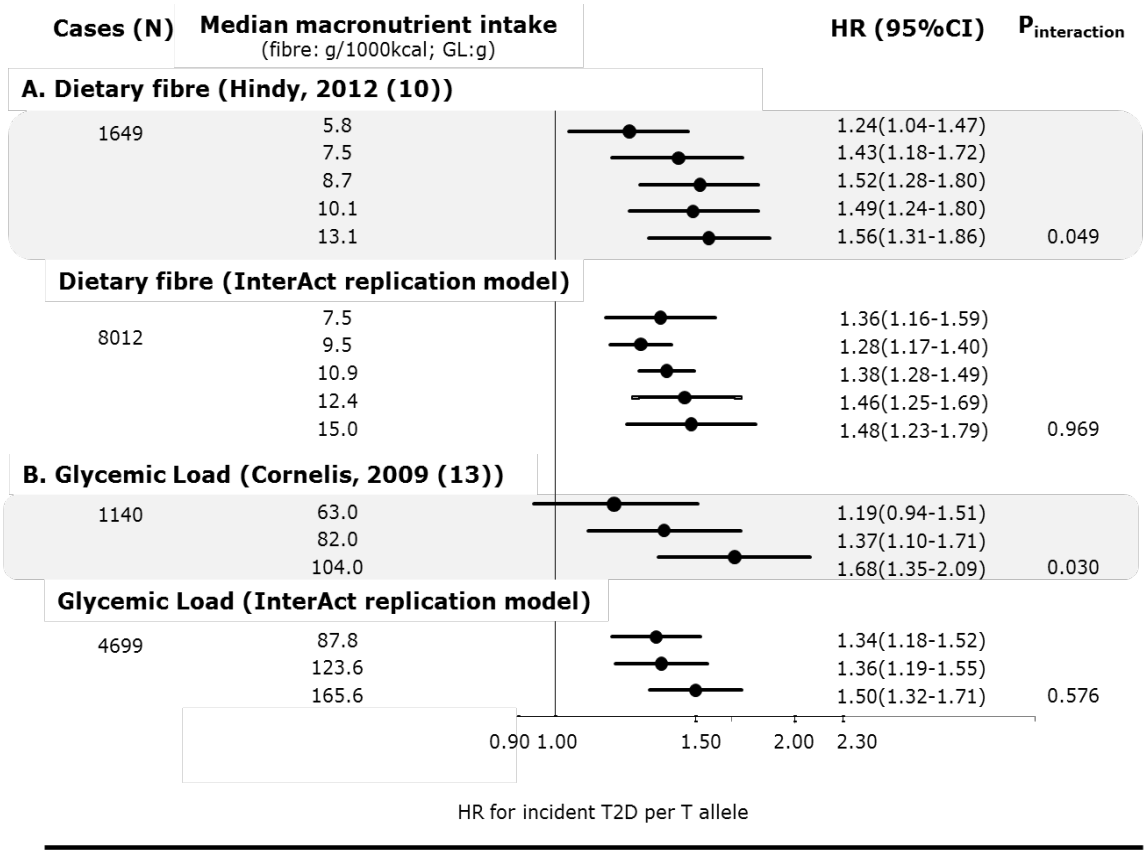



**Figure 3-2: Interaction between genetic variants within *TCF7L2* and dietary fibre or glycaemic load: comparison between the study by Hindy et al., 2012 and Cornelis et al., 2009 with EPIC-InterAct**

A) Odds ratio (OR) from Hindy et al., 2012[4] (above) and pooled hazard ratios (HR) from EPIC-InterAct (below) for type 2 diabetes (T2D) per T allele of rs7903146 (*TCF7L2*) and quintiles of dietary fibre (g/1000kcal). Hindy et al., adjusted for age, sex, body-mass index (BMI), total energy intake, season and method (dietary intake assessment method); EPIC-InterAct replication model adjusted for age (=underlying time scale), sex, centre, BMI, total energy intake, and season, excluding the EPIC-InterAct centre: Malmo.

B) OR from Cornelis et al., 2009[172] and HR from EPIC-InterAct for T2D per T allele of rs12255372 (*TCF7L2*) by tertiles of glycaemic load (g). In women only. Cornelis et al. adjusted for age, BMI, smoking, alcohol, coffee, menopausal status, physical activity, energy-adjusted ratio of PUFA/SFA and trans fat and cereal fibre; EPIC-InterAct adjusted for age (=underlying time scale), centre, BMI, smoking, alcohol, coffee, menopausal status, physical activity, energy-adjusted ratio of PUFA/SFA and cereal fibre. Given that Cornelis et al., evaluated this interaction in a female cohort (Nurses' Health Study), the EPIC-InterAct analysis was conducted in women only.

$P_{interaction}$ for EPIC-InterAct: estimated by treating macronutrient and SNPs as continuous variables. In EPIC-InterAct, heterogeneity between countries was not significant ($I^2$=0% in A and 1% in B).

These two SNPs (rs7903146 and rs12255372) are in moderate linkage disequilibrium (CEU $r^2$=0.7).

Please see methods for an explanation of the range of country-specific dietary intake percentiles used.

Total sample size for EPIC-InterAct analysis of the interaction between dietary fibre and *TCF7L2* interaction: 18292; GL and *TCF7L2*: 11992 (women only).

Statistical test: multiplicative interaction analysis using Prentice-weighted Cox regression

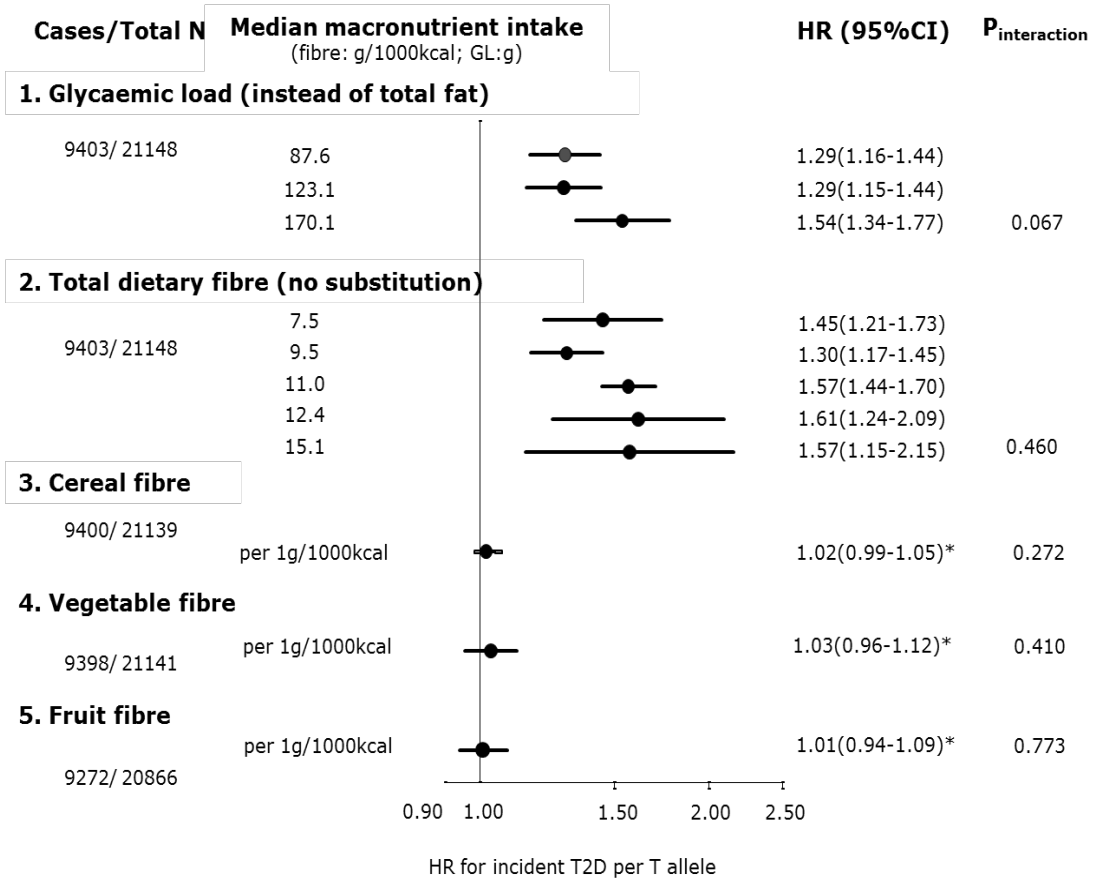



**Figure 3-3: Interaction between genetic variants within *TCF7L2* and dietary fibre or glycaemic load under the modified model with EPIC-InterAct**

**1. EPIC-InterAct meta-analysis of the hazard ratio (HR) of T2D per T allele of rs12255372 (*TCF7L2*), stratified by percentiles of glycaemic load (g)** with adjustment for age (=underlying time scale), sex, centre, eigenvectors (first 5 principal components for population stratification), physical activity, education, smoking, sex-specific alcohol categories, total energy intake, season of intake assessment, dietary intake (protein, magnesium, iron, vitamin C, coffee, tea, leafy vegetables and artificially sweetened beverage) and BMI. Heterogeneity between countries was moderate ($I^2$=44%).

**2. EPIC-InterAct meta-analysis of the HR of T2D per T allele of rs7903146 (*TCF7L2*), stratified by percentiles of dietary fibre intake (g/1000kcal)** with adjustment for age (=underlying time scale), sex, centre, eigenvectors (first 5 principal components for population stratification), physical activity, education, smoking, sex-specific alcohol categories, total energy intake, season of intake assessment, dietary intake (carbohydrate, protein, saturated fat, mono-unsaturated fat, poly-unsaturated fat, magnesium, iron, vitamin C, coffee, tea, leafy vegetables and artificially sweetened beverage) and BMI. Heterogeneity between countries was not significant ($I^2$=13%).

**3-5. EPIC-InterAct meta-analysis of the relative hazard ratios (RHR) for incident T2D per increase in T allele (TCF7L2 rs7903146) and per 1g/1000kcal of dietary fibre. By fibre sub-types.**

with adjustment for age (=underlying time scale), sex, centre, eigenvectors (first 5 principal components for population stratification), physical activity, education, smoking, sex-specific alcohol categories, total energy intake, season, dietary factors (carbohydrate, protein, SFA, MUFA, PUFA, magnesium, iron, vitamin C, coffee, tea, leafy vegetables and artificially sweetened beverage) and BMI. All models for different subtypes of fibre are mutually adjusted. Heterogeneity $I^2$ between countries differed for vegetable fibre x*TCF7L2*: 51%; fruit fibre x*TCF7L2*: 71% and cereal fibre x*TCF7L2*: 0%

$P_{interaction}$: estimated by treating macronutrient and SNPs as continuous variables

*relative hazard ratio

Statistical test: multiplicative interaction analysis using Prentice-weighted Cox regression

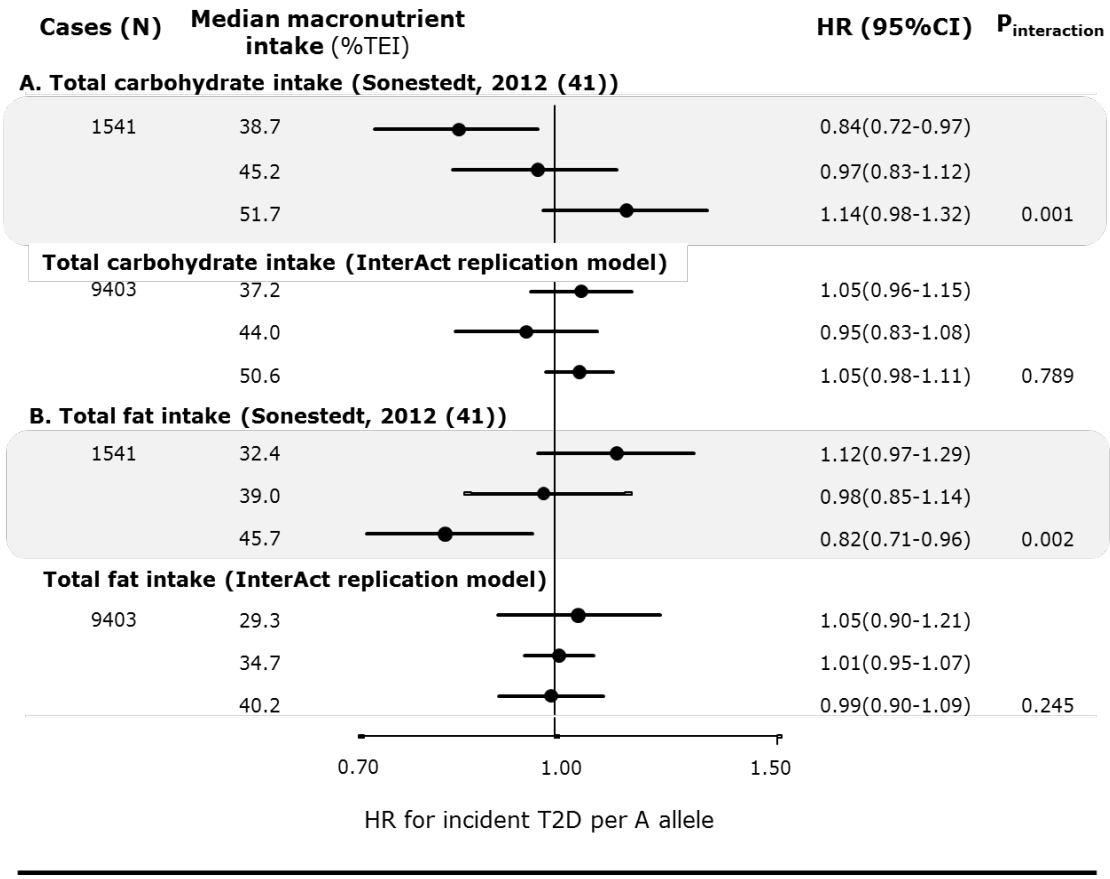



**Figure 3-4: Hazard ratio (HR) of incident T2D per A allele of rs10423928 (*GIPR*) by tertiles of macronutrient intake: comparison between the study by Sonestedt et al., 2012 and EPIC-InterAct.**

For both A) total carbohydrate intake and B) total fat intake, HR from Sonestedt et al., 2012[194] (above) and pooled HR from EPIC-InterAct (below). Sonestedt et al., adjusted for age, sex, physical activity, education, smoking, sex-specific alcohol categories, season, total energy intake (TEI), method and BMI. EPIC-InterAct replication adjusted for age (=underlying time scale), sex, centre, physical activity, education, smoking, sex-specific alcohol categories, season, total energy intake and BMI

$P_{interaction}$ for EPIC-InterAct: estimated by treating macronutrients and rs10423928 as continuous variables.

In EPIC-InterAct, heterogeneity between countries was not significant ($I^2$=17% in A; 19% in B)

Please see methods for an explanation of the range of country-specific macronutrient intake percentiles used.

Total sample size for EPIC-InterAct analysis: 21148.

Statistical test: multiplicative interaction analysis using Prentice-weighted Cox regression

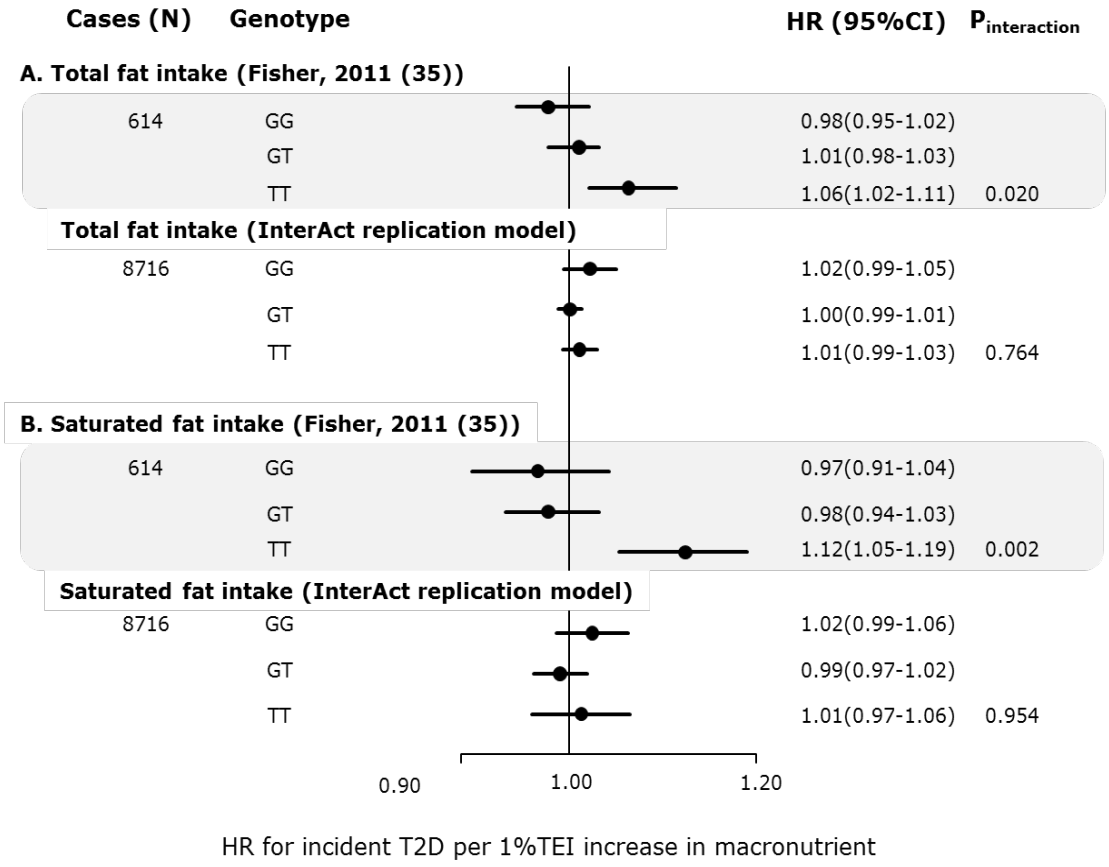



**Figure 3-5: Hazard ratio (HR) of incident T2D per 1% total energy intake (TEI) increase in macronutrient intake, stratified by *CAV2* rs2270188 genotype: comparison between the study by Fisher et al., 2011 and EPIC-InterAct.**
For both A) total fat intake and B) saturated fat intake, HR from Fisher et al.,[186] (above) and pooled HR from EPIC-InterAct (below). Fisher et al., adjusted for sex, age, total energy intake, and BMI ($p_{interaction}$ using results from the confirmatory case-cohort study under the additive genetic model). EPIC-Interact replication model adjusted for age (=underlying time scale), sex, centre, total energy intake, BMI, excluding the EPIC-InterAct centre: Potsdam.
Note: the classical interaction model was adopted, not the genotype-specific model reported in Fisher et al., because of the stated equivalence of the 2.
$P_{interaction}$: estimated by treating macronutrients and rs2270188 as continuous variables.
In EPIC-InterAct, heterogeneity between countries was moderate ($I^2$ =41% in A; 34% in B).
Please see methods for an explanation of the range of country-specific macronutrient intake percentiles used.
Total sample size for EPIC-InterAct analysis: 19477.
Statistical test: multiplicative interaction analysis using Prentice-weighted Cox regression

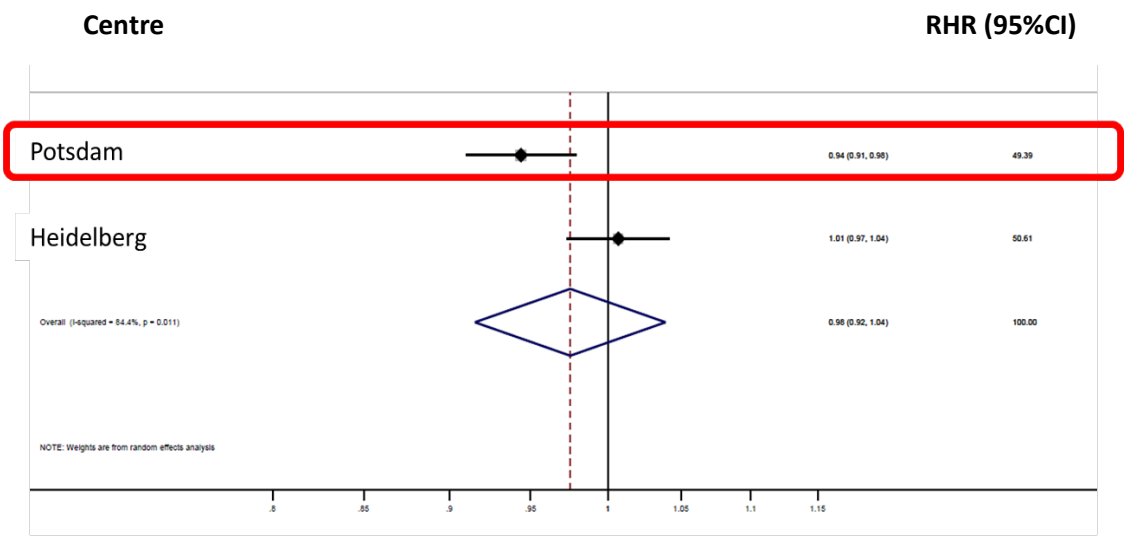



**Figure 3-6: Relative Hazard Ratio (RHR) for incident T2D per 1% total energy increase in total fat intake and per T allele of rs2270188 (*CAV2*) in 2 German centres within EPIC-InterAct.**

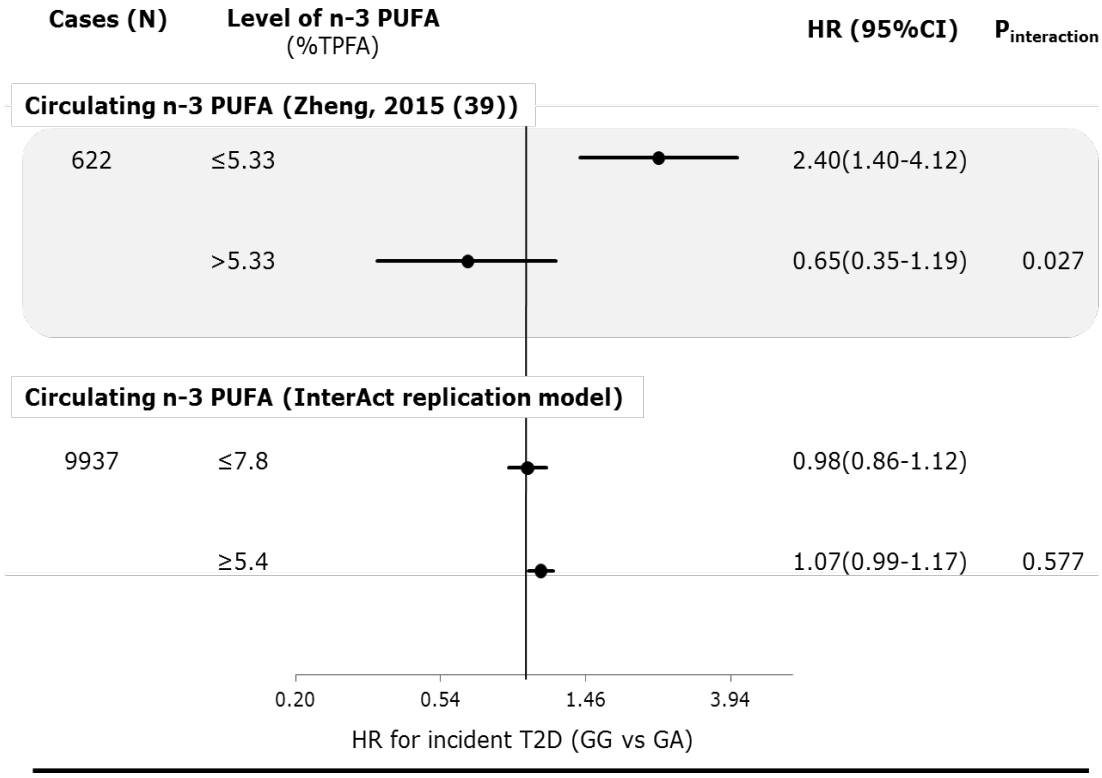



**Figure 3-7: Interaction between genotypes for rs3786897 (PEPD: GA vs GG) and percentage of total phospholipid fatty acid (TPFA) that is circulating n-3 PUFA: comparison between the study by Zheng et al., 2015 and EPIC-InterAct.**

Odds ratio (OR) from Zheng et al., 2015[192] (above) and pooled hazard ratio (HR) from EPIC-InterAct (below) for T2D. Zheng et al., adjusted for age and sex. EPIC-InterAct replication model adjusted for age (=underlying time scale), sex, centre.

$P_{interaction}$: estimated by treating circulating n-3 PUFA as dichotomous and *PEPD* rs3786897 as continuous variables.

In EPIC-InterAct, heterogeneity between countries was not significant ($I^2$ =15%).

Please see methods for an explanation of the range of country-specific circulating n-3 PUFA percentiles used.

Abbreviation: TPFA: total phospholipid fatty acid.

Total sample size for EPIC-InterAct analysis: 22273.

Statistical test: multiplicative interaction analysis using Prentice-weighted Cox regression

There was also no evidence of any significant interaction in the more detailed analysis that accounted for additional potential confounders and isocaloric macronutrient substitution (Appendix D, Figures 1 to 3).

For the interaction between dietary fibre and *TCF7L2* on incident T2D, there was no material difference between the analysis using calibrated and uncalibrated dietary fibre (Figure 4-8).

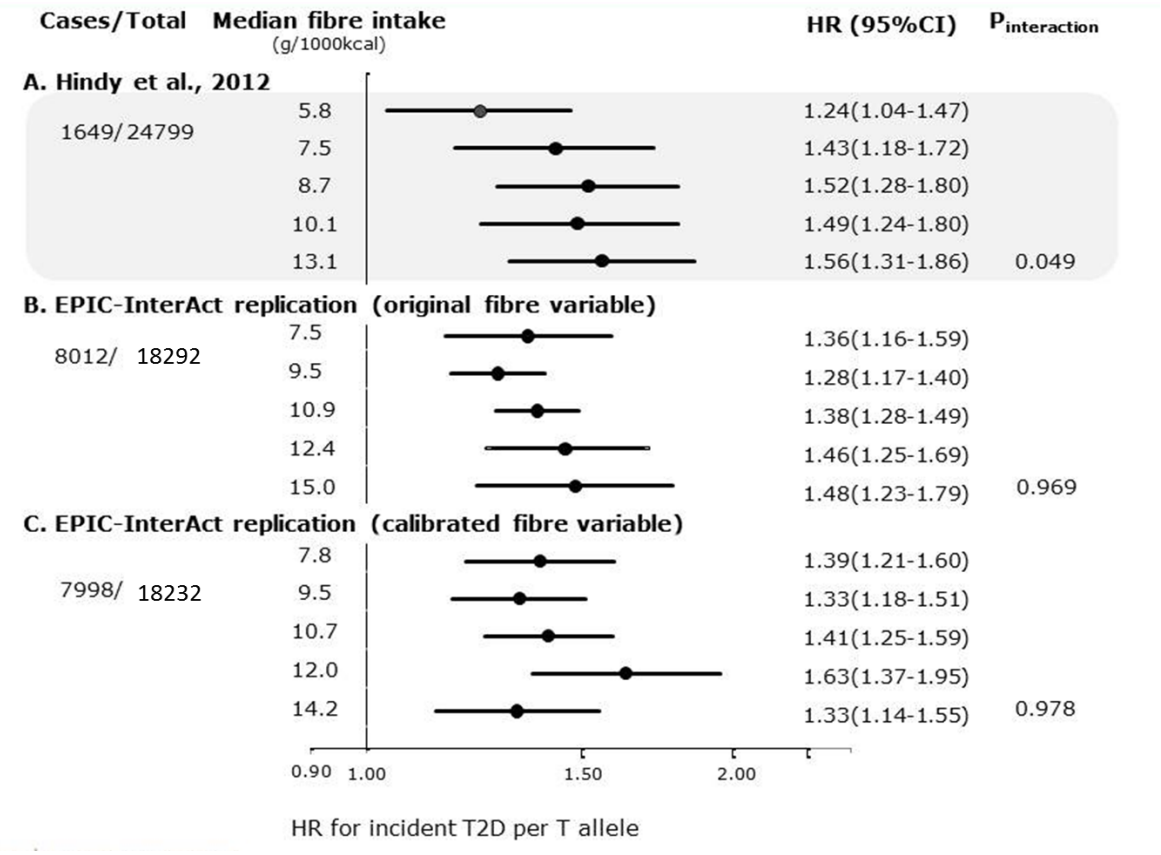



**Figure 3-8 Interaction between genetic variants within *TCF7L2* and dietary fibre: comparison between analysis by Hindy et al., analysis in EPIC-InterAct with original fibre variable (uncalibrated) and analysis in EPIC-InterAct with calibrated fibre variable.**

A) Odds ratio (OR) from Hindy et al., 2012[4], B) pooled hazard ratios (HR) from EPIC-InterAct (below) for type 2 diabetes (T2D) per T allele of rs7903146 (*TCF7L2*) and quintiles of dietary fibre (g/1000kcal) using uncalibrated dietary fibre and C) pooled HR from EPIC-InterAct for T2D per T allele of rs7903146 (*TCF7L2*) and quintiles of dietary fibre (g/1000kcal) using calibrated dietary fibre. Hindy et al., adjusted for age, sex, body-mass index (BMI), total energy intake, season and method (dietary intake assessment method); EPIC-InterAct replication model adjusted for age (=underlying time scale), sex, centre, BMI, total energy intake, and season, excluding the EPIC-InterAct centre: Malmo.

$P_{interaction}$ for EPIC-InterAct: estimated by treating macronutrient and SNPs as continuous variables.

Statistical test: multiplicative interaction analysis using Prentice-weighted Cox regression

## *4.5 Discussion*

13 papers reporting gene-macronutrient interactions on T2D were identified from this systematic review but we did not find any consistently replicated evidence for gene-macronutrient interaction in the aetiology of T2D. Next, given the nature of these findings, a general discussion about the shared issues which arose from this review, rather than a discussion about specific interaction findings, will be presented.

### Challenges in identifying and replicating gene-macronutrient interactions

The differences between findings from the published studies and EPIC-InterAct re-emphasise the challenges in studying gene-diet interactions. We believe that selective reporting through limited consideration for multiple testing in studies examining multiple SNPs and/or macronutrients, without a justified predefined hypothesis and lack of replication are among several possible methodological explanations for this inconsistency. There are also other factors that may explain why we find different results to that of the published studies, which include heterogeneity in dietary measurement, population under study, study design or in analysis and reporting, as discussed in previous reviews.[119,142,199,200]

In hypothesis-free genetic epidemiological analyses, given the large number of variants tested on a genome-wide scale, stringent correction for multiple testing has attempted to minimise the false positive rate.[201] However, approaches to address multiple testing for interaction studies have been less consistent. We found in our review that studies often used a nominal $p<0.05$ as the threshold for rejecting the null, even when performing many tests,[4,170,172,186,194] including one study that performed an exploratory analysis of 256 gene-macronutrient interactions and used $p<0.05$ for rejecting the null.[186] Two of the six studies that reported significant interactions would have passed multiple testing corrections after Bonferroni correction,[170,194] whereas only one study adopted Bonferroni corrected p values.[192] Therefore, we consider false positive reports as a potential explanation for the discordant findings between EPIC-InterAct

and published reports. Whilst debate continues about whether an optimal p-value threshold should exist for interaction studies,[202] in the future researchers should account for potential inflation of a false positive rate when conducting multiple interaction analyses (e.g. by using methods such as the 'effective number of independent tests'),[203] preferably with independent internal replication in additional studies. As evidenced by genome-wide association studies, the design of genetic studies allows for relatively straightforward *in silico* replication, yet few gene-macronutrient interaction studies that report finding a significant interaction have been followed by independent replication.[4,170,172,192,194] Arguably, variations in dietary assessment methods introduce more difficulty in identifying suitable replication sources. For instance, although four independent studies included in our review examined the interaction between *TCF7L2* and dietary fibre or related fibre subtypes,[4,170,172,189] it is arguable how comparable their methods are. This includes differences in relation to dietary assessment and degree of measurement bias: two studies used an FFQ,[170,172] one used a 24-hour recall[189] and one used a combined FFQ, diet history and 7-day diary.[4] Moreover regarding study design: two studies were prospective,[4,170] and two were cross-sectional,[172,189] which may be subject to differing levels of bias and ability to determine the direction of effect. Lastly, analytic methods varied by whether variables were treated as continuous or categorical and what covariates were controlled for. Therefore, in the future, researchers should consider internally conducted replication as this would reduce variation in the analysis.

We attempted to mirror the population and analyses conducted in EPIC-InterAct with that of the published studies reporting an interaction and showed comparable characteristics except with one study where the ethnicity was different (an Asian population was examined).[192] However, we cannot exclude possible heterogeneity between studies. This may include differences in study design (only one published study used a case-cohort study design similar to EPIC-InterAct) and unmeasurable inconsistencies in dietary exposures (e.g. food composition, preparation methods, measurement tool used, coding of exposures) between countries within EPIC-InterAct and between EPIC-InterAct and the published studies. Indeed, this was evident for the interaction between

*CAV2* with total fat and SFA, which showed centre-specificity. Within the German centres, an interaction was detected for the EPIC-InterAct centre Potsdam but not Heidelberg resulting in an overall lack of interaction for Germany (Figure 4-6 shows the interaction results for fat and *CAV2* by German centres). However, the percentage of total variation due to heterogeneity across the countries within EPIC-InterAct was low to moderate for interactions under the replication model ($I^2$: 14 to 30%). The consistently null findings across different countries of EPIC-InterAct strengthen the inference from this overall null finding. Another possible contributor to the disparity between results, for instance relating to *TCF7L2* and dietary fibre, may be overestimation by certain estimation parameters (e.g. odds ratios) which could lead to an inflated difference between fibre categories.[204]

The methodological issues described above highlight difficulties in discerning whether type I error or true heterogeneity underlies the inconsistencies we observed and are similar to that faced in the broader gene-environmental literature.[35,202] For gene-environment interactions, recommendations have been made for improving standards in design, analysis and reporting, which are also relevant for gene-diet studies.[141,142] For example, Cornelis suggested minimising publication bias by publishing both positive and negative interaction findings and reporting them in supplemental materials if necessary.[142]

### Strengths and limitations

A potential limitation of our systematic review is that the heterogeneity between the published studies did not enable a quantitative synthesis (e.g. meta-analysis) or formal statistical evaluation of publication bias, as previously demonstrated and advised against by Palla et al.,[154] We did, however, use a comprehensive search strategy and tried to minimise publication bias by contacting authors of studies possibly examining interactions (n=4).

From the literature, EPIC-InterAct, is the largest study of incident T2D cases (>5 times the number of previous studies) with both genetic data and measures of self-reported macronutrient intake as well as objective circulating fatty acids, which overall makes it well positioned to examine these reported interactions (power calculations in the Table 4-5 below). The prospective design minimises

the potential bias due to reverse causality for dietary exposures. Additionally, to our knowledge, this is the first study for gene-macronutrient interactions which has investigated the effect of isocaloric macronutrient substitution in the observational setting. This is important for public health interpretation of macronutrient density if total daily energy intake is fixed, as the benefit of decreasing one macronutrient may be dependent on which macronutrient/s it is replaced by. As exemplified by the interaction between rs10423928 (*GIPR*) and both fat and carbohydrate intake.[194] The authors did not model isocaloric macronutrient substitution, so it is unclear whether they demonstrated a lower risk of T2D among those possessing an A allele for this genetic variant and consuming a diet that is higher in fat and lower in carbohydrate intake reflects a substitution effect or independent effects for the respective macronutrient.

Several limitations need to be considered whilst interpreting the results. These analyses only investigated a select number of interactions which have been reported in the literature. Hence, this does not preclude the possibility that there may be interactions between other dietary (including foods and dietary patterns) and/or other genetic variants or combined gene scores. Moreover, our focus was on examining possible type I error. Given that we did not examine interactions which did not reach statistical significance in published studies (possible type II error) we cannot preclude the presence of genuine interactions among those loci we did not test. Alternative study designs, such as agnostic GEWIS may be better placed to investigate the presence of these interactions.[205] Variations in dietary assessment between EPIC-InterAct centres may contribute to potential variation in measurement bias for macronutrients. The current literature consists of studies primarily from European populations, which limits the generalisability of our findings. Another limitation is the uncertainty about between-centre and between-country variation in the degree of measurement bias for macronutrient intake and how this may affect our interaction results, especially for dietary fibre as previously explained. Hence, the interaction between dietary fibre and *TCF7L2* using a variable calibrated with a single measure of 24-hour recall,[188] was examined and found no difference to interaction results using the uncalibrated dietary fibre variable. This can be explained by two reasons, i) Hoffmann and colleagues cautions that the accuracy of calibrated FFQs depends

on the accuracy of the reference method.[188] Although a single 24 hour recall measured in <8% of EPIC participants has been validated against 24 hour urine nitrogen as a reference method for calibrating protein intake,[187] it is unclear how reliable and accurate it would be as the reference method for calibrating dietary fibre intake when objective measurements are as yet unavailable for dietary fibre. ii) Yet, even if a more accurate reference was available for dietary fibre intake since regression calibration only affects the precision of effects, when the precision of effects in both strata of a dietary exposure under an interaction model is improved, it is unlikely that the significance of interaction estimates would be affected. Moreover, no previous interaction studies have calibrated their macronutrient intake data. There are alternative suggestions to address measurement bias within an interaction framework proposed by Midthune and colleagues,[206] but this is beyond the scope of this thesis.

**Table 3-5: Power calculations**

| **Published study** | **Interactions reported (effect allele)** | **Effect allele frequencies (Europeans)** | **Marginal effect in InterAct (gene/macronutrient)** All log-additive unless specified | **LogOR for interaction** (highest /lowest category of intake or genotype) | **Power to detect interaction in InterAct** |
|---|---|---|---|---|---|
| Hindy et al., 2012 | *TCF7L2*-dietary fibre (rs7903146) (T) | 0.23 | 1.3/1 | 1.26 | 0.87 |
| Wirstrom et al., 2012 | *TCF7L2*-cereal fibre (rs7903146) (T) | 0.23 | 1.3/0.97 | 1.72 | 0.99 |
| Cornelis et al., 2009 | *TCF7L2*-GL (rs12255372) (T) | 0.21 | 1.3/0.997 | 1.41 | 0.99 |
| Sonestedt et al., 2012 | *GIPR*-CHO rs10423928 (A) | 0.17 | 1/0.98 | 1.36 | 0.96 |
| | *GIPR*-fat | 0.17 | 1/1 | 0.73 | 0.92 |
| Fisher et al., 2011 | *CAV2*- total fat rs2270188 (T) | 0.55 | 1/1 | 1.01 | 0.05 |
| | *CAV2*- SFA | 0.45 | 1/1 | 1.15 | 0.58 |
| Zheng et al., 2015 | *PEPD*- circulating n-3 PUFA rs3786897(A) | 0.59 | 1/0.88 Dominant model | 0.29 | 0.99 |

Assumptions for QUANTO (v.1.2.4): InterAct sample size= 9,000 cases, 12,000 non-cases; Prevalence of T2D: 10%; macronutrient intake normally distributed; $\alpha$ <0.05; per increase in 1%E from macronutrient intake per risk allele, under a log additive model unless specified otherwise

### Implications for public health and research

Our systematic review and replication in EPIC-InterAct highlight the importance of independent replication in the study of interaction and the need to improve standards in conducting and reporting future interactions. Moreover, our review reveals a gap in non-candidate gene approaches to examining gene-macronutrient interactions. This includes GRS and GEWIS. Additionally, there is a gap for examining gene-macronutrient interactions in the intervention setting. Given that no promising gene-macronutrient interactions were found and that genetic variants most relevant for interactions may be those with weak or no marginal effects,[95,207] GEWIS may aid in discovering novel interactions at potentially unexpected genetic loci. Furthermore, this review highlights that on the basis of the interactions examined here, there is no evidence to support genetic personalisation of macronutrient intake recommendations as a strategy to prevent T2D.

### Recommendations for future research

A) Within study considerations: (i) specifying the hypothesis of the study and accounting for multiple testing, as appropriate; (ii) reporting all interaction results and if conducted as pre-planned or post-hoc, regardless of whether findings are positive, negative or null; and (iii) where possible, notable interaction findings should be accompanied with independent replication and if not feasible a discussion should be made about the reasons and validity of non-replicated findings.

B) General considerations for studies within the field: (i) improving consistency and standards in examining and reporting interactions, as advised in several previous reviews; [141–143] (ii) conducting studies examining non-European populations; and (iii) applying isocaloric macronutrient substitution.

## *4.6 Conclusion*

Whilst there is growing interest in personalised diets to more effectively combat cardiometabolic conditions such as T2D, none of the gene-macronutrient

interactions currently reported in the literature could be replicated in the large-scale EPIC-InterAct study. We also do not find evidence to support that the specific gene-macronutrient interactions we examined play a significant role in the aetiology of T2D. Improving standards in examining and reporting interactions, including independent replication, will be vital to making progress in this area.

# *Chapter 5 Interaction between genetic risk score and macronutrient intake on incident T2D in EPIC-InterAct*

This chapter follows on from the previous chapter by addressing some of the major research gaps identified from our systematic review. In particular, it contributes to the currently scarce literature on gene-macronutrient interactions and T2D, using genetic risk scores. This project relates to objective 3 (Chapter 2).

This Chapter will soon be published:

**Li SX**, Imamura F, Ye Z, Schulze MB, Zheng J, Ardanaz E, et al., Interplay between genetic predisposition and macronutrient intake on type 2 diabetes incidence: analysis within EPIC-InterAct across eight European countries. *Diabetologia* 2018. doi: 10.1007/s00125-018-4586-2. [Epub ahead of print]

## *5.1 Abstract*

**Objective** Gene-macronutrient interactions may contribute to the development of type 2 diabetes (T2D) but evidence is inconclusive.

**Research design and methods** the interactions between three genetic risk scores (GRS) and macronutrient intake on the development of T2D was examined in EPIC-InterAct, a prospective case-cohort study across eight European countries (N=21,900, 9,742 incident T2D cases). GRS were constructed for T2D (48 single nucleotide polymorphisms [SNPs]), insulin resistance (53 SNPs) and body mass index (97 SNPs) based upon shared biology in T2D development. Macronutrient intake was estimated from diets reported in questionnaires, including the proportion of energy derived from total carbohydrate, protein, fat, plant and animal protein, saturated, mono- and poly-unsaturated fat and dietary fibre. Using multivariable-adjusted Cox regression, country-specific interaction results were estimated under the multiplicative scale, which were pooled by random-effects meta-analysis. Secondary analysis accounted for isocaloric macronutrient substitution.

**Results** A significant positive association was observed between total and animal protein intake with incident T2D. No interactions were identified between any of the three GRS and any macronutrient intake, with low to moderate heterogeneity between countries ($I^2$ range: 0-51.6%). Results were similar when analyses accounted for isocaloric macronutrient substitution, when using weighted and unweighted GRS and when examining individual SNPs.

**Conclusions** Genetic susceptibility to T2D, insulin resistance and BMI did not modify the association between macronutrient intake and incident T2D. This indicates that dietary recommendations for macronutrient intake for preventing T2D apply across the population regardless of genetic predisposition to these three metabolic conditions.

## *5.2 Background*

An emphasis on macronutrient composition has dominated public health dietary recommendations for decades, with guidance on optimal percent of energy to be consumed from carbohydrate, fat and protein. In addition to guidance relating to macronutrient quantity, more recently dietary guidance has evolved to acknowledge the importance of macronutrient quality. For instance, evidence supporting cardiometabolic benefits when replacing dietary saturated fat with polyunsaturated fat has led to emergence of guidance concerning fat subtype or quality.[81,208] Similarly, there is evidence on the importance of carbohydrate quality such as higher dietary fibre intake,[86] which is reflected in guidance on preventing T2D.[10]

Genetic susceptibility to T2D is also well documented, with heritability estimated to be between 40% and 80% and meta-analyses of genome-wide association efforts have now identified over 70 genetic variants associated with increased risk of T2D, that explain around 6% of its variance.[25,27] With this, there has been increasing interest in whether this genetic susceptibility may differentially influence how macronutrient intake affects the development of T2D (gene-macronutrient interaction) and whether this may support the rising popularity of 'personalised' or 'precision' nutrition.[95] However, evidence for this from a recent systematic review for interactions between genetic predisposition and macronutrient intake for T2D is far from conclusive.[209]

Genetic risk scores (GRS), composed of multiple single nucleotide variants, are of interest because they may better reflect the polygenic nature of T2D.[125] It is known that insulin resistance (IR), often secondary to obesity, is involved in the aetiology of T2D and susceptibility genetic loci have been identified for T2D, IR and Body Mass Index (BMI) [5,28,30]. Moreover, GRS for T2D, IR, and BMI may help to explain more variance for T2D than candidate genetic variant approaches and may improve statistical power to detect potential interactions. Yet, there is a paucity of studies examining gene-macronutrient interaction using a GRS approach.

We aimed to increase understanding about the aetiology of T2D by investigating potential interactions between genes and macronutrient intake on the incidence of T2D using GRS for each of these three traits: T2D, IR and BMI.

## *5.3 Methods*

The EPIC-InterAct cohort, genotyping and imputation, dietary assessment and estimation of macronutrient intake and statistical analyses have already been described in the Methods Chapter (Chapter 3). In addition to these general methods, below I describe some specific methods for this piece of work.

*Cohort*

This current study is based on 21,900 adults with available genome-wide genotyping and dietary data (9,742 cases and 12,158 non-cases) from EPIC-InterAct.

*Computation of genetic risk score*

Unweighted GRS for T2D, IR and BMI were generated by summing up the number of risk alleles for each trait [125]. SNPs from loci reaching genome-wide significance for the respective traits in published meta-analyses investigating European populations were used (48 for T2D after excluding BMI raising alleles *FTO* and *MC4R* so as to reduce overlap with the BMI GRS; 53 for IR and 97 SNPs for BMI).[5,28,30] An additive genetic model was assumed for each SNP. Taking the T2D GRS (48 SNPs) as an example, an individual could carry 0, 1, or 2 risk alleles for each genotyped SNP or a continuous dosage for imputed SNPs (the possible range for T2D GRS: 0 to 96). The results of these analyses were also compared with interactions examined using weighted GRS based on the beta coefficients from published meta-analyses of GWAS for the respective traits (available for BMI and T2D).[5,28]

*Imputation of missing covariates*

In situations where data is missing at random but not completely at random, conducting complete case analyses may be biased. Additionally, where there may be different degrees of missingness across several covariates, this may result in a loss of precision and power. Since the difference between missing at random or completely at random is difficult to disentangle statistically, covariates at <30% missing (Table 5-1) were imputed using Multiple Imputations by Chained Equations (MICE) in *Stata*.[210] In brief, practically this method involved the inclusion of covariates needing imputation as well as the outcome and predictors of all missing covariates. Firstly, all continuous covariates and predictors were normalised to enable prediction based upon a known distribution. Based on the predictors in the model MICE uses a Bayesian approach to create a specified number of datasets in addition to the original, with the plausible imputed values in place of those that were previously missing within each dataset. 20 datasets are usually sufficient to allow for uncertainty in imputation so that the datasets would capture the supposed 'real' value within the variation of imputed values. The method explained by White and colleagues involved running the analysis model for each imputed dataset independently after which results from each analysis would be then pooled into one estimate using Rubin's rules. However, for each of the interaction analyses in EPIC-InterAct, this was conducted by country and genotyping chip before being meta-analysed. At the time of conducting this analysis, there was no function within MICE to incorporate this additional complexity. Therefore, for computational efficiency, a single imputed dataset was used for all analyses, after confirming no obvious between-imputation variation across twenty multiple-imputation datasets[210] (Figure 5-1 provides an example analyses across all 20 datasets). The total variation, being the sum of the within dataset variation (3.89E-05) plus the between-study variation (1.32E-07) was very small. Therefore, all datasets were deemed comparable. Dataset number 15 was therefore chosen, at random, for all analyses presented in this paper. With many imputation methods, there are limitations related to MICE, including the possibility of misleading results driven by systematic differences in missing values. Therefore complete-case

analysis was also performed as a sensitivity analysis to compare with the results from MICE.[211]

**Table 4-1: Imputed baseline variables from EPIC-InterAct**

| Characteristics | % missing | Imputed |
|---|---|---|
| **Number** | | 21900 |
| **Age (y)** | - | |
| **Sex (%male)** | - | |
| **PA level (%)** | 1.1% | Y |
| inactive | | |
| moderately inactive | | |
| moderately active | | |
| active | | |
| **Highest school level (%)** | 1.7% | Y |
| none | | |
| primary school | | |
| technical/professional | | |
| secondary school | | |
| longer education (inc. university) | | |
| **BMI (kg/m$^2$)** | 0.7% | Y |
| **Waist hip ratio** | 7.9% | Y |
| **Smoking status (%)** | 1% | Y |
| never | | |
| former | | |
| current smoker | | |

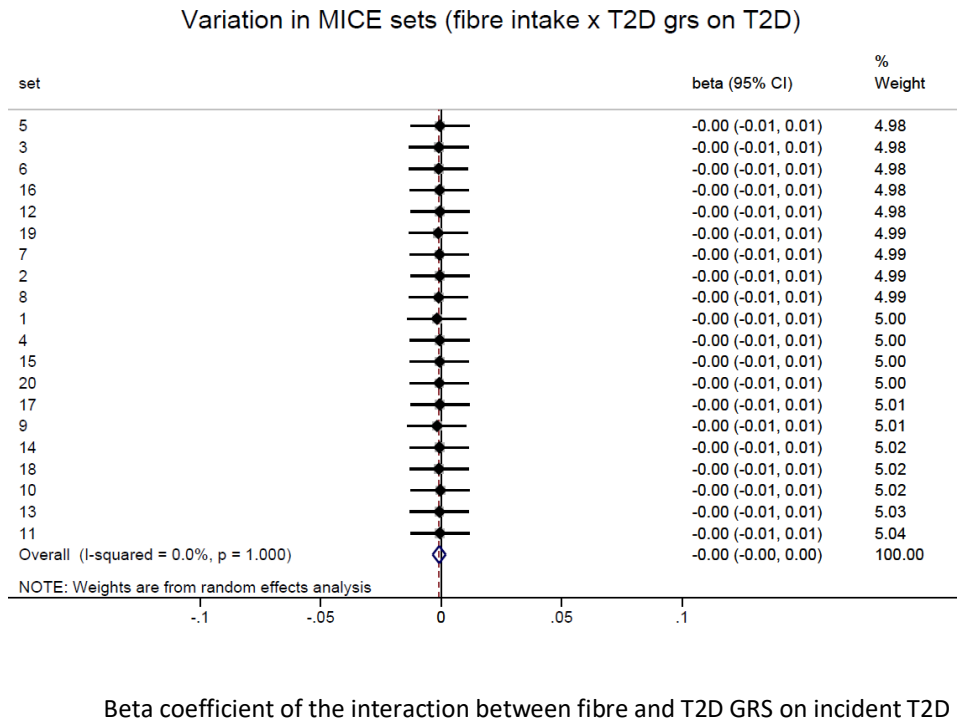



**Figure 4-1: Comparison between 20 imputed datasets for the interaction between total fibre intake and T2D genetic risk score on incident T2D.**
Beta co-efficient for the interaction term adjusted for age (=underlying time scale), sex, centre, TEI, eigenvectors (first 5 PC for population stratification), physical activity, education, smoking, sex-specific alcohol categories, carbohydrate, SFA, MUFA, PUFA intake, magnesium, iron, vitamin C, leafy vegetable, tea, coffee, BMI

## *Statistical analyses*

### *Main effect analyses*

The associations between macronutrient intake and T2D were estimated by treating macronutrient exposures as continuous variables (per SD difference in percentage of total energy intake). Potential outliers were Winsorised at the 1st and 99th percentile.

Crude and multivariable-adjusted Prentice-weighted Cox regression models were constructed within country. For consistency, modelling was based as closely as possible on those used in previous EPIC-InterAct analyses for carbohydrate,[212] protein[213] and dietary fibre.[86] For dietary fat and subtypes (not previously published in EPIC-InterAct), models adjusted for age (underlying time scale), sex, centre (nominal categorical), total energy intake (Kcal/d), physical activity (inactive, moderately inactive, moderately active, active), education (none, primary school, technical/professional, secondary school, longer education (including university)), smoking (never, former, current smoker), sex-specific alcohol intake (none, light drinking: 0.1-6g/d,

moderate drinking: men 6.1-24g/d and women 6.1-12g/d, heavy drinking: men>24g/d and women>12g/d) and other dietary confounders (dietary fibre, magnesium, iron, vitamin C, green leafy vegetables, tea and coffee in mg or g/day). The covariates used can be found in the legend of Table 5-2. Country-specific hazard ratios (HR) for each macronutrient intake were combined across countries using random-effects meta-analysis.

For the association between GRS and T2D, the GRS were treated both as continuous (per SD difference) and dichotomised exposures (as high and low GRS estimates based on being above or below the median estimates among those in the subcohort) (Table 5-3). Prentice-weighted Cox regression models were constructed within country and by genotyping chip. Genotype chip-specific and then country-specific estimates were combined using random-effects meta-analysis. Analyses were adjusted for age (underlying time scale), sex, centre, the first 5 principal components for population stratification and BMI.

*Interaction analyses*

For the interaction analyses between each macronutrient intake and each GRS on the risk of developing T2D, both exposures were treated as continuous variables (GRS per SD difference and macronutrient as densities, being 5% of total energy intake/day and 1g/1000kcal/day for dietary fibre) to avoid loss of statistical power from categorisation. Multiplicative interaction was evaluated by fitting a product term between the GRS and macronutrient exposures [135]. Regression models were constructed in the same way as in analysis of the association between GRS and incident T2D, and the list of covariates included in the models were the same as for the main associations between macronutrient intake and T2D (described above), with addition of the first 5 principal components for population stratification. Between-country heterogeneity was quantified by the $I^2$ value and *P* for heterogeneity was derived from the Cochran-Q test. BMI was a covariate in the interaction analysis for IR and T2D GRS only.

For visualisation, the estimated HR for each dietary factor was stratified by high and low GRS groups (Figure 5-2).

Given that GRS may mask interactions with individual SNPs, further secondary interaction analysis was conducted for each SNP within all 3 GRSs. The potential effect of whether substitution of a macronutrient for another was modified by genetic predisposition while energy intake was held constant (i.e. isocaloric macronutrient substitution using the multivariate nutrient density model),[168] was also examined. This was performed for energy-bearing macronutrients (not dietary fibre), with the modelling strategy provided in Appendix F.

*Stata* v14 (StataCorp LP, Texas, USA) was used for analysis. Numerical p value for interaction were reported in tables and figures, however, the threshold for determining statistical significance for interactions between GRS and macronutrient intake (without isocaloric macronutrient substitution) was ≤0.0015 (0.05/33 tests) to account for the effective number of independent tests among correlated exposures (see Table 5-4 for correlations).[203] The threshold for determining statistical significance for interactions under the substitution model was ≤0.0006 (0.05/81 tests).

## *5.4 Results*

At baseline, the mean age of participants in the subcohort was 52.3 years (SD= 9.3 years) and median follow-up was 10.9 years. A statistically significant association was observed between proportion of energy from protein intake and incident T2D (HR per 1 SD =1.10; 95%CI: 1.03, 1.18) and from animal protein intake and incident T2D (HR per 1 SD =1.10; 95%CI: 1.01, 1.18) (Table 5-2). Those with higher genetic risk for T2D, IR or BMI (i.e. in the 'high GRS' group) were at a significantly higher risk for developing T2D than those in the lower genetic risk group (Table 5-3). There was little correlation between the GRS and macronutrients (Table 5-4).

**Table 4-2: Association between macronutrient intake and the incidence of type 2 diabetes: EPIC-InterAct study**

| | No. Cases/Total | Mean (SD) | Mean (SD) | HR (95%CI) per SD* |
|---|---|---|---|---|
| | | Subcohort non-cases | Total incident T2D cases | |
| **Median years of follow up** | 9742/21900 | 12.3 | 6.8 | |
| **Age at baseline(y)** | | 52.3 (9.3) | 55.7 (7.6) | |
| **Sex (%male)** | | 37.9% | 49.9% | |
| **Macronutrient intake** | | | | |
| Carbohydrate (%TEI) | 9742/21900 | 44.1 (6.9) | 43.7 (6.9) | 0.97 (0.92,1.02) |
| Protein (%TEI) | 9742/21900 | 16.9 (3.0) | 17.2 (3.0) | 1.10 (1.03,1.18) |
| Animal protein (%TEI) | 9742/21900 | 10.5 (3.2) | 10.9 (3.2) | 1.10 (1.01,1.18) |
| Plant protein (%TEI) | 9742/21900 | 5.00 (1.3) | 4.9 (1.3) | 1.074 (0.999,1.150) |
| Fat (%TEI) | 9742/21900 | 34.8 (5.7) | 34.7 (5.7) | 1.03 (0.99,1.08) |
| Saturated fat (%TEI) | 9742/21900 | 13.4 (3.3) | 13.3 (3.3) | 0.99 (0.93,1.06) |
| Monounsaturated fat (%TEI) | 9742/21900 | 13.1 (3.4) | 13.0 (3.4) | 1.04 (0.97,1.12) |
| Polyunsaturated fat (%TEI) | 9742/21900 | 5.5 (1.8) | 5.6 (1.8) | 1.066 (0.999,1.137) |
| Fibre (g) | 9742/21900 | 22.7 (7.5) | 22.6 (7.6) | 0.92 (0.84,1.02) |
| Fibre-cereal (g) | 9739/21891 | 8.8 (4.9) | 8.9 (4.9) | 0.96 (0.86,1.07) |
| Fibre-fruit (g) | 9608/21611 | 4.3 (3.2) | 4.2 (3.2) | 0.86 (0.73,1.02) |
| Fibre-vegetable (g) | 9737/21893 | 4.1 (2.6) | 34.0 (2.6) | 0.99 (0.94,1.04) |

Hazard Ratios (HR) for macronutrients (per SD) and incident Type 2 Diabetes (T2D): Carbohydrate intake adjusted for age (=underlying time scale), sex, centre, education, physical activity, smoking status, sex-specific alcohol categories, BMI, total energy intake, dietary protein, PUFA:SFA ratio, dietary fibre (attempt to replicate model 3 in Sluijs et al., 2013); Protein intake and subtypes adjusted for age (=underlying time scale), sex, centre, physical activity, smoking status, sex-specific alcohol categories, BMI, waist hip ratio, total energy intake, dietary fibre, SFA, MUFA, PUFA, soft drinks, tea, and coffee (not adjusted for carbohydrates; i.e., a substitution model), education (attempt to replicate model 4 in van Nielen et al., 2014); Fat intake and subtypes adjusted for age (=underlying time scale), sex, centre, physical activity, smoking status, sex-specific alcohol categories, BMI, total energy intake, dietary fibre, magnesium, iron, vitamin C, leafy vegetables, tea, coffee, education; Dietary fibre and subtypes adjusted for age (=underlying time scale), sex, centre, physical activity, smoking status, sex-specific alcohol categories, total energy intake, dietary carbohydrates, magnesium, saturated fatty acids, education level. Fibre subtypes were mutually adjusted (attempt to replicate model 3 in Aune et al., 2015).
*SD calculated based on the whole population

**Table 4-3: Association between genetic risk scores for BMI, IR and T2D and the incidence of type 2 diabetes: EPIC-InterAct Study**

| Unweighted genetic risk score | N | HR (95%CI) | |
|---|---|---|---|
| | | Model 1 | Model 2 |
| T2D | | | |
| per SD (4.3 risk alleles) | | 1.42(1.32,1.52) | 1.49(1.37,1.63) |
| High (≥52 risk alleles) | 12033 | 1.75(1.55,1.98) | 1.85(1.58,2.16) |
| Low (<52) | 9867 | 1 [Reference] | 1 [Reference] |
| IR | | | |
| per SD (4.5 risk alleles) | | 1.11(1.08,1.14) | 1.14(1.09,1.20) |
| High (≥55) | 11303 | 1.18(1.12,1.25) | 1.23(1.12,1.35) |
| Low (<55) | 10597 | 1 [Reference] | 1 [Reference] |
| BMI | | | |
| per SD (6.3 risk alleles) | | 1.07(1.04,1.10) | NA |
| High (≥91) | 11230 | 1.12(1.06,1.19) | |
| Low (<91) | 10670 | 1 [Reference] | |

Hazard Ratio (HR) for genetic risk scores and Type 2 Diabetes (T2D):
Model 1: age (=underlying time scale), sex, centre, first 5 principal components (PC) for population stratification
Model 2: model 1 + BMI
Abbreviations: T2D- type 2 diabetes, BMI- body mass index, IR- insulin resistance, SD- standard deviation
Number of SNPs: T2D (48 as per Morris et al., Nature Gen, 2012), BMI (97 as per Locke et al., Nature, 2015), IR (53 as per Lotta et al., Nature Gen, 2016)

## Interaction between genetic risk scores and macronutrient intake on incident T2D

The association between the proportion of energy derived from the intake of each macronutrient and incident T2D did not differ significantly by T2D GRS (lowest $p_{interaction}$=0.20) (Figure 5.2 and Table 5.5). No significant interactions were found with GRS for IR (lowest $p_{interaction}$ =0.21) or BMI (lowest $p_{interaction}$=0.22) (Figure 5.2 and Table 5.5). There was low to moderate heterogeneity between countries in EPIC-InterAct ($I^2$ range: 0-51.6%) (Table 5.5). Results did not change substantially when GRS were weighted by the beta coefficient of each SNP on their respective trait from published GWAS meta-analyses (data not presented).

Similarly, no interactions were detected when modelling isocaloric macronutrient substitution for the GRS-based analyses ($p_{interaction}$ ≥0.17) (Appendix F, model 5).

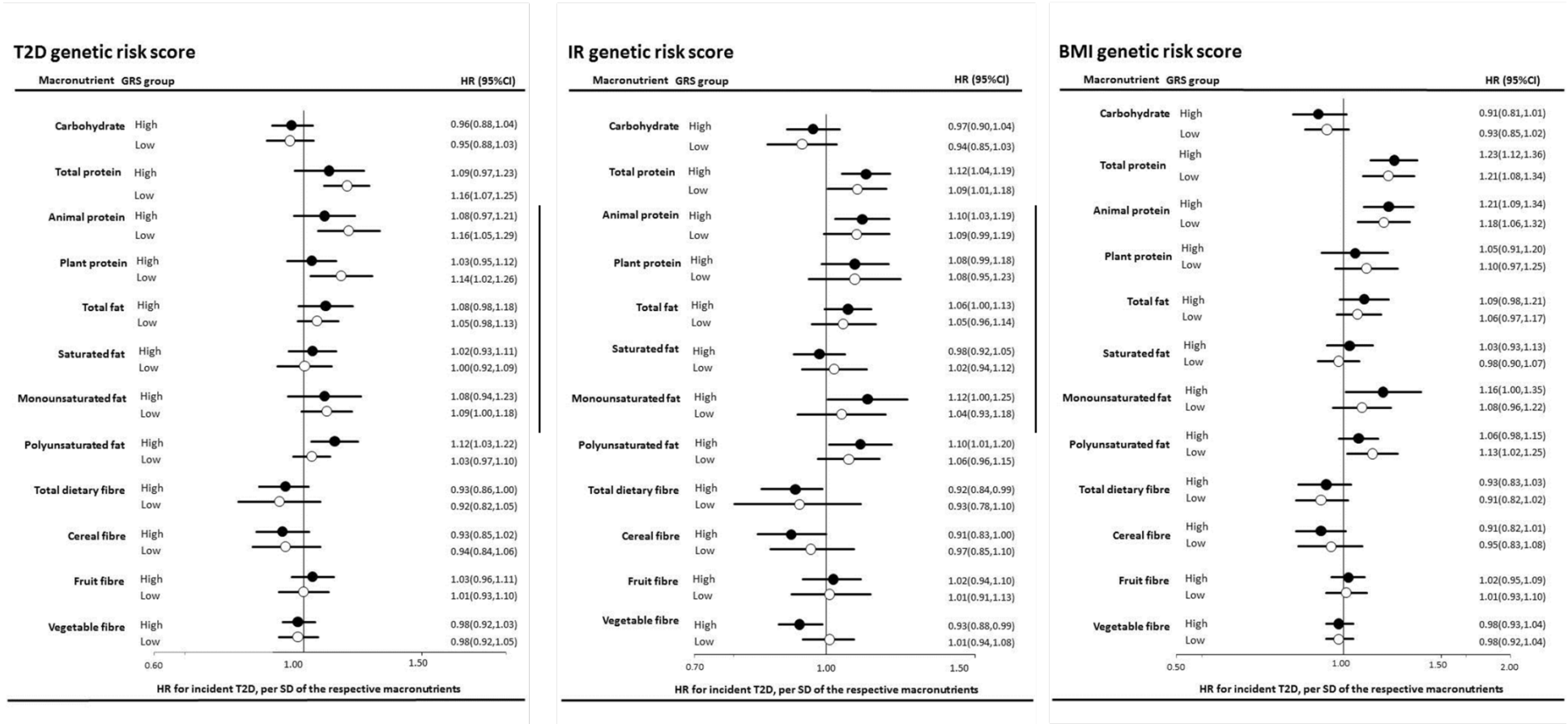



**Figure 4-2: Association between macronutrient intake (per SD) and the incidence of type 2 diabetes, which were stratified by high or low genetic risk scores: EPIC-InterAct Study**
Macronutrients are modelled per standard deviation difference in intake (see Table 5-1 for the standard deviation for each macronutrient)
Carbohydrate intake adjusted for age (=underlying time scale), sex, centre, education, physical activity, smoking status, sex-specific alcohol categories, BMI, total energy intake, dietary protein, PUFA:SFA ratio, dietary fibre and first 5 principal components (PC) for population stratification; Protein intake and subtypes adjusted for age (=underlying time scale), sex, centre, physical activity, smoking status, sex-specific alcohol categories, BMI, waist hip ratio, total energy intake, dietary fibre, SFA, MUFA, PUFA, soft drinks, tea, and coffee (not

adjusted for carbohydrates; i.e., a substitution model), education and first 5 principal components (PC) for population stratification; Fat intake and subtypes adjusted for age (=underlying time scale), sex, centre, physical activity, smoking status, sex-specific alcohol categories, BMI, total energy intake, dietary fibre, magnesium, iron, vitamin C, leafy vegetables, tea, coffee, education and first 5 principal components (PC) for population stratification; Dietary fibre and subtypes adjusted for age (=underlying time scale), sex, centre, physical activity, smoking status, sex-specific alcohol categories, total energy intake, dietary carbohydrates, magnesium, saturated fatty acids, education level and first 5 principal components (PC) for population stratification. Fibre subtypes were mutually adjusted.
Interaction analysis for BMI GRS does not adjust for BMI.
Interactions were considered statistically significant if $p < 0.0015$ (0.05/33 tests)
There were also no statistically significant multiplicative interactions between any macronutrient and GRS, when treating them as continuous exposures (see Table 5.5).
Filled in black dots= high genetic risk score, Unfilled white dots= low genetic risk score
Example of interpretation: the HR of 1 SD difference in fruit fibre on incident T2D is 1.03 in those who have the highest genetic predisposition for T2D and 1.01 for those with lower genetic predisposition for T2D. There was no statistically significant difference between those with different genetic predispositions for T2D.

**Table 4-4: Correlations between dietary exposures within the EPIC-InterAct Study**

| Exposures | T2D GRS | IR GRS | BMI GRS | fat | sfa | mufa | pufa | prot | prota | protp | cho | fb | fb_veg | fb_fruit | fb_cereal |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **fat** | 0.01 | 0.01 | 0.00 | | 0.71 | 0.70 | 0.40 | 0.01 | 0.13 | 0.37 | 0.61 | 0.30 | 0.03 | 0.20 | 0.29 |
| **sfa** | 0.01 | 0.01 | 0.01 | | | 0.21 | 0.11 | 0.21 | 0.04 | 0.57 | 0.31 | 0.35 | 0.27 | 0.30 | 0.05 |
| **mufa** | 0.00 | 0.00 | 0.00 | | | | 0.03 | 0.15 | 0.21 | 0.12 | 0.52 | 0.26 | 0.03 | 0.01 | 0.41 |
| **pufa** | 0.00 | 0.01 | 0.02 | | | | | 0.01 | 0.01 | 0.03 | 0.25 | 0.06 | 0.15 | 0.11 | 0.00 |
| **prot** | 0.00 | 0.00 | 0.03 | | | | | | 0.04 | 0.04 | 0.32 | 0.17 | 0.37 | 0.18 | 0.15 |
| **prota** | 0.00 | 0.00 | 0.02 | | | | | | | 0.25 | 0.47 | 0.05 | 0.24 | 0.11 | 0.25 |
| **protp** | 0.01 | 0.00 | 0.01 | | | | | | | | 0.30 | 0.63 | 0.36 | 0.24 | 0.36 |
| **cho** | 0.01 | 0.00 | 0.02 | | | | | | | | | 0.37 | 0.07 | 0.24 | 0.35 |
| **fb** | 0.01 | 0.00 | 0.01 | | | | | | | | | | 0.53 | 0.52 | 0.43 |
| **fb_veg** | 0.01 | 0.01 | 0.01 | | | | | | | | | | | 0.32 | 0.13 |
| **fb_fruit** | 0.02 | 0.00 | 0.01 | | | | | | | | | | | | 0.13 |
| **fb_cereal** | 0.01 | 0.00 | 0.00 | | | | | | | | | | | | |

**Strength of correlation**

| |
|---|
| 0 |
| 0.3 |
| 0.5 |
| 0.7 |
| 1 |

Spearman's Rho, without direction of correlation.
Abbreviations: sfa- saturated fatty acid, mufa- monounsaturated fatty acid, pufa- polyunsaturated fatty acid, prota- animal protein, protp- plant protein, cho- carbohydrate, fb- fibre, fb_veg: vegetable fibre, fb_fruit: fruit fibre, fb_cereal: cereal fibre.
All variables treated as continuous variables
Correlations calculated based on subcohort population only (N=12749)
Effective number of independent tests:[203] for the interaction between genetic risk scores and macronutrient interactions (without isocaloric macronutrient substitution): 10.9776 (variance of the observed eigenvalues: 1.12). This estimates the independent number of tests, accounting for any correlated macronutrients.

**Table 4-5: Multiplicative interaction between macronutrient and unweighted genetic risk scores: EPIC-InterAct Study**

| Macronutrient intake (5% total energy intake) | Model | GRS for body mass index (per 6.3 risk alleles)[%] | | | GRS for insulin resistance (per 4.5 risk alleles) | | | GRS for type 2 diabetes (per 4.3 risk alleles) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Beta (95% CI) | *P* | $I^2$ (%) | Beta (95% CI) | *P* | $I^2$ (%) | Beta (95% CI) | *P* | $I^2$ (%) |
| **Carbohydrate** | *Sluijs et al., 2013* | 0.001(-0.034,0.035) | 0.971 | 50.2 | 0.001(-0.03,0.031) | 0.970 | 19.1 | 0(-0.028,0.027) | 0.976 | 0.0 |
| **Total protein** | *van Nielen et al., 2014* | -0.032(-0.1,0.036) | 0.351 | 0.0 | 0.062(-0.035,0.16) | 0.210 | 37.9 | -0.06(-0.162,0.042) | 0.249 | 43.4 |
| **Animal protein** | *van Nielen et al., 2014* | -0.022(-0.083,0.039) | 0.475 | 0.0 | 0.055(-0.048,0.158) | 0.293 | 51.6 | -0.05(-0.145,0.045) | 0.302 | 45.5 |
| **Plant protein** | *van Nielen et al., 2014* | -0.057(-0.206,0.091) | 0.451 | 0.0 | 0.004(-0.241,0.249) | 0.975 | 49.0 | -0.022(-0.185,0.142) | 0.793 | 0.0 |
| **Total fat** | *model 4* | 0.01(-0.017,0.036) | 0.471 | 0.0 | -0.004(-0.048,0.041) | 0.869 | 41.2 | 0.011(-0.021,0.043) | 0.497 | 0.0 |
| **SFA** | *model 4* | 0.031(-0.019,0.08) | 0.228 | 0.0 | -0.015(-0.089,0.058) | 0.687 | 22.8 | 0.027(-0.034,0.089) | 0.384 | 0.0 |
| **MUFA** | *model 4* | 0.014(-0.044,0.073) | 0.634 | 0.0 | 0.023(-0.087,0.133) | 0.685 | 48.5 | -0.016(-0.097,0.065) | 0.699 | 15.3 |
| **PUFA** | *model 4* | -0.042(-0.155,0.072) | 0.471 | 30.5 | 0.01(-0.139,0.16) | 0.895 | 40.4 | 0.072(-0.039,0.182) | 0.204 | 2.2 |
| **Total dietary fibre (g/1000kcal)** | *Aune et al., 2015* | 0.004(-0.006,0.015) | 0.408 | 0.0 | -0.004(-0.019,0.011) | 0.623 | 44.7 | -0.001(-0.011,0.01) | 0.919 | 0.0 |
| **Vegetable fibre (g/1000kcal)** | *Aune et al., 2015* | 0(-0.029,0.028) | 0.973 | 6.8 | -0.012(-0.053,0.028) | 0.548 | 46.4 | 0.007(-0.021,0.035) | 0.634 | 0.0 |
| **Fruit fibre (g/1000kcal)** | *Aune et al., 2015* | 0.013(-0.008,0.033) | 0.222 | 0.0 | -0.007(-0.036,0.022) | 0.632 | 40.8 | -0.001(-0.029,0.026) | 0.927 | 30.2 |
| **Cereal fibre (g/1000kcal)** | *Aune et al., 2015* | -0.005(-0.022,0.012) | 0.568 | 0.0 | -0.002(-0.02,0.015) | 0.799 | 0.0 | 0.004(-0.018,0.025) | 0.735 | 21.5 |

Abbreviations: SFA- saturated fatty acid, MUFA- monounsaturated fatty acid, PUFA- polyunsaturated fatty acid, GRS- genetic risk score

**Beta-coefficient for the interaction between each of the genetic risk score and respective macronutrients on incident T2D, adjusted for the following covariates (as per previously published EPIC-InterAct study, see methods).**

**Carbohydrate** age (=underlying time scale), sex, centre, education, physical activity, BMI, smoking status, sex-specific alcohol categories, total energy intake, dietary protein, PUFA:SFA ratio, dietary fibre, first 5 principal components (PC) for population stratification

**Protein and subtypes**

age (=underlying time scale), total energy intake (TEI), centre, and sex, smoking, education, physical activity, sex-specific alcohol categories, dietary fibre, SFA, MUFA, PUFA, soft drinks, tea, and coffee (not adjusted for carbohydrates; i.e., a substitution model), BMI, waist hip ratio, first 5 PC for population stratification

**Fat and subtypes**

Model 1: age (=underlying time scale), sex, centre, TEI, first 5 PC for population stratification

Model 2: model 1 + lifestyle factors- physical activity, education, smoking, sex-specific alcohol categories
Model 3: model 2+ dietary covariates (dietary fibre, magnesium, iron, vitamin C, leafy vegetables, tea, coffee)
Model 4: model 3+ BMI

**Dietary fibre and subtypes**

age (=underlying time scale), sex, smoking status, physical activity, education level and sex-specific alcohol categories, total energy intake, dietary carbohydrates, magnesium, saturated fatty acids, first 5 PC for population stratification, types of fibre were mutually adjusted
Trying to replicate model 3.
[&] interactions with BMI GRS does not adjust for BMI
IR GRS and protein intake interaction does not adjust for centre because of convergence issues
**Example of interpretation:** the beta-coefficient of the interaction between total fat and BMI GRS was 0.010 for incident T2D.

In the analysis examining interactions with each individual SNP (N tests=5,265) and isocaloric macronutrient substitution, none of the interactions tested was statistically significant (Figure 5-3 and Appendix G).

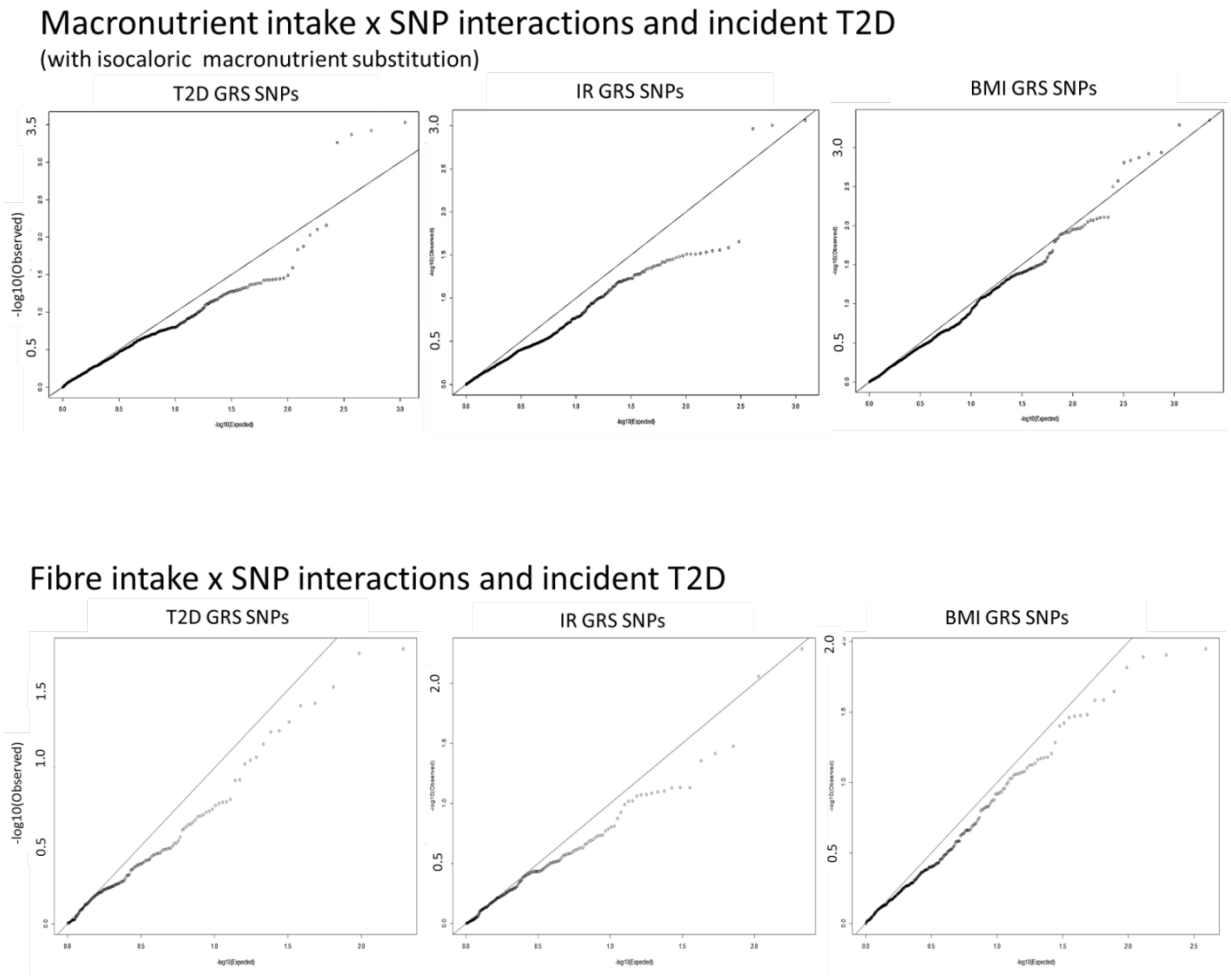



**Figure 4-3: QQ plot of pvalues for the interactions between individual SNPs for each of the genetic risk scores and macronutrients or dietary fibre on T2D (with isocaloric macronutrient substitution): EPIC-InterAct Study**
All interactions are based on the most adjusted models (previously reported in Appendix F). None of the individual SNP interactions were significant after accounting for multiple testing. P value for interaction threshold for significance <9.50E-6 (0.05/5265 tests).
Please note that each individual GRS graph contains multiple sets of interaction analyses, therefore the expected p values may be conservative due to correlated analyses. That is, for the interaction between fibre intake and T2D GRS, this includes interactions between T2D GRS and total fibre, T2D GRS and cereal fibre, T2D GRS and vegetable fibre and T2D GRS and fruit fibre on incident T2D
A list of SNP x macronutrient interactions with pvalue for interaction<0.05 are available in Appendix G (Excel)

Results were similar between the current analysis based on imputed data and a complete case analysis (9,403 cases and 11,745 non cases) (Table 5-6 provides an example).

**Table 4-6: An example of interaction findings comparing multiple imputation and complete case analysis**

| Unweighted T2D GRS (per SD) x Macronutrient (g/1000kcal) | Multiple imputation analysis results (9742 cases, 12158 noncases) | | Complete case analysis results (9403 cases, 11745 noncases) | |
|---|---|---|---|---|
| | **Beta (se)** | ***P*** | **Beta (se)** | ***P*** |
| *Total dietary fibre* | 0.001(0.006) | 0.875 | 0.001(0.007) | 0.839 |
| *Cereal fibre* | 0.006(0.011) | 0.606 | 0.002(0.016) | 0.902 |

Abbreviation: SSB: sugar sweetened beverage, BMI: body mass index, T2D: type 2 diabetes, GRS: genetic risk score, p: p value for interaction

Modelling the same as that in Table 5.5, model 4.

**Conclusion: no substantial difference in results between analysis approaches**

## *5.5 Discussion*

In this large, multi-country, population-based study from Europe, no statistically significant interactions were observed between three metabolic GRS and macronutrient intake on the development of T2D. In other words, based on these findings, the association between macronutrient intake and incident T2D does not seem to vary by genetic susceptibility for T2D, IR or BMI. The public health implication for prevention of T2D of these null findings is that currently dietary guidelines on macronutrient intake should apply universally across populations that may differ in genetic susceptibility to metabolic indices.[10]

The literature on gene-macronutrient interaction studies and T2D is limited, with a previous cross-sectional study which examined the interaction between a T2D GRS and carbohydrate and fibre intake failing to identify interactions for prevalent T2D (N= 1,337 cases of T2D).[189] This current analysis is the first to examine gene-macronutrient interaction for T2D risk prospectively, investigating major macronutrients including quantity and quality indicators of fat, protein

and carbohydrate consumption, using a T2D GRS with additional genetic variants (48 compared to 15 in the previous work), as well as also examining GRS for BMI and IR in a much larger sample (with 9,742 T2D cases). There is an inconsistent evidence base for intermediate end points related to T2D (e.g. adiposity and glycaemic traits).[214–217] The studies which reported an interaction were limited by small sample size (N<715).[216,217] In particular, a study which recruited users of a nutrigenetic service and reported interactions between dietary fat intake and a GRS composed of obesity and lipid metabolism SNPs on adiposity risk [216] is limited by both its cross-sectional design and its likely potential for gene-diet dependence. It is unclear whether the GRS used in that study may be correlated with fat related dietary exposures given that it included SNPs involved in lipid metabolism (e.g. the lipoprotein lipase gene, *LPL,* which hydrolyses triglycerides and mediates lipoprotein uptake into cells). Dudbridge and Fletcher posited that gene-environment dependence may result in spurious interactions.[218] That is, if an environmental exposure mediates the association between a genetic exposure and the outcome, testing the interaction of that environmental and genetic exposure on the outcome may lead to a spurious interaction because it is more likely that the outcome would occur when both exposures are present. The evaluation of the likelihood of such dependence and of possible type I error can be facilitated by examining the correlation between the GRS and macronutrient intakes, as have been performed in Table 5-4. In this current study, the consistency across various methods (adoption of unweighted and weighted GRS and a combined GRS as well as single SNPs) collectively strengthens the confidence in the null findings for interaction that were observed.

The null findings may indicate that interactions between the three GRS that were examined and macronutrient intake may truly not exist for T2D, or if they do potentially exist, the current approach was not able to detect them. There are several possible reasons that may contribute to the absence of interactions in this study. From a dietary perspective, foods and/or dietary patterns, may offer greater insights than nutrients based on the food synergism hypothesis.[219] This proposes that biological synergism may occur within the 'food matrix' as a whole, including with yet potentially unidentified constituents. Indeed interactions have been reported between the GRS for BMI and intake of sugar sweetened beverages

on risk of developing obesity ($p_{interaction}$ <0.001).[148,220] Furthermore, foods or dietary patterns may be subject to less accumulated measurement bias (i.e. bias from self-report of food intake plus error from conversion to nutrients). From a genetics perspective there may be other genetic loci, with no or weak marginal genetic effects for the traits of interest that may show a significant variation in effect between subgroups of the population. Therefore, future agnostic approaches such as genome-environment-wide-interaction studies, are of interest. Additionally, it may be valuable to examine specific candidate genetic variants when biologically plausible hypotheses of specific interactions may exist.[209,221] Another possible explanation for null interaction findings may be that in certain situations many factors must be present simultaneously for an interaction of interest to become apparent.[222] Multi-way interactions have been reported but these were not tested in this current study so future research is warranted.[195]

Among strengths, EPIC-InterAct is a large population based study with incident T2D and both genetic data and harmonised measures of self-reported macronutrient intake. All three GRS were positively associated with incident T2D (Table 5-2) [5,28,30] and the associations between macronutrient intake and T2D were directionally consistent with previous literature (Table 5-1).[86,212,213,223,224] The prospective design minimises the potential bias due to recall bias and reverse causality for dietary exposures and the verification of diabetes cases minimises possible misclassification bias of the outcome. To our knowledge, this study represents the most comprehensive investigation of the interaction between multiple GRS and macronutrient intake on incident T2D, to date. An attempt was made to address some of the key methodological issues identified from our recent systematic review, including multiple testing and inadequate control for likely confounders.[209] To reduce the risk of spurious gene-macronutrient interactions, it was confirmed that GRS were not correlated with macronutrient intake.[218] To our knowledge this is also the first observational study for gene-macronutrient interactions within the cardiometabolic literature that has investigated the effect of isocaloric macronutrient substitution. This is important for public health interpretation of macronutrient density, as the benefit of decreasing one macronutrient may be dependent on which macronutrient(s) it is replaced by.

Apart from issues of generalisability and the limited dietary and genetic exposures that were examined, as discussed above, other limitations merit consideration. Despite attempts to reduce confounding, including by controlling for population stratification in addition to other known risk factors for T2D,[126] the possibility of residual confounding cannot be fully excluded. Given that interactions were examined using continuous exposures, the possibility that they exist for specific categories of either exposure, cannot be fully excluded. However, using continuous exposures aimed to maximise statistical efficiency. The generalisability of these findings is limited to European populations and research is warranted in different populations.

In conclusion, within a multi-centre European cohort, we observed no interaction between GRSs for type 2 diabetes, insulin resistance and BMI and macronutrient intake on the risk for developing type 2 diabetes. These findings suggest that currently there is no support for personalised dietary advice on macronutrient intake for type 2 diabetes prevention in subgroups of the population defined by their overall genetic risk for type 2 diabetes, insulin resistance or BMI.

# *Chapter 6 Interaction between genetic risk score and food or beverage intake on incident T2D in EPIC-InterAct*

This chapter examines the interaction between genetic risk scores and foods and beverages in the development of T2D. It relates to objective 4 (Chapter 2).

## *6.1 Abstract*

**Background** There has been inconsistent evidence about whether the association between foods and beverages with type 2 diabetes (T2D) might differ depending on an individual's genetic predisposition for metabolic disorders. We aimed to investigate the interaction between foods and beverages and genetic susceptibility on the risk of developing T2D.

**Methods and Findings** The current analyses included 9,742 incident T2D cases and 12,158 non-cases in EPIC-InterAct, a prospective case-cohort study across eight European countries. Genetic risk scores (GRS) for body mass index (97 single nucleotide polymorphisms: SNPs), insulin resistance (53 SNPs) and T2D (48 SNPs) were chosen based upon shared biology in T2D development. Fifteen dietary variables that were previously examined with T2D within published meta-analyses were identified: fruits, green leafy vegetables, root

vegetables, wholegrain breads and cereals, legumes, nuts and seeds, fermented dairy, red meat, processed red meat, fish, egg and egg products, sugar sweetened beverage (SSB), coffee and tea. Using multivariable-adjusted Cox regression, we estimated country-specific interaction results using multiplicative and additive scales which were pooled by random-effects meta-analysis. Processed red meat and SSB intake were positively associated, whereas coffee and tea intake were inversely associated with T2D. After accounting for multiple testing (45 tests, p value threshold for significance <0.001), no significant interactions were identified between any of the GRS and foods or beverages on incident T2D (p for multiplicative interactions≥0.01; for additive interactions≥0.01).

**Conclusions** Genetic susceptibility for BMI, insulin resistance or T2D did not significantly modify the association between food or beverage intake and T2D incidence and therefore does not support dietary recommendations stratified by metabolic genetic risk for the prevention of type 2 diabetes.

## *6.2 Introduction*

Current nutritional guidelines emphasise consumption of, for example, green leafy vegetables and wholegrain foods and a reduction in red and processed red meats and sugar-sweetened beverages (SSB) for the prevention of T2D.[10,41]

The effect of these foods on metabolic risk may be heterogeneous by genotype, arguing a case for the presence of interaction between dietary factors and genes. For instance, previous research suggests that the adverse consequences of habitual SSB intake on adiposity and the risk of developing obesity, a strong determinant for T2D, were significantly higher in those with a high compared to those with a low genetic predisposition to obesity.[148,220] Whether the observed interaction between SSB and BMI genetic risk score (GRS) is also relevant in the aetiology of T2D is currently unknown. If genes contribute to the heterogeneity in dietary response, then it is important to consider the potential for personalised dietary advice based on genotype for preventing T2D.[3] Studies have reported, for instance, that consuming wholegrain bread and cereals and coffee influence T2D risk differently depending on an individual's *TCF7L2* gene.[170,171,221] However, there is a gap in knowledge about whether there are interactions for other foods and beverages that may be important in the development of T2D (e.g. green leafy vegetables, red meat, SSB, among others).

The advantages of using a GRS has been previously discussed (Chapter 5). However, to date, there has been limited evidence for gene-food interactions using GRS.[221,225]

We, therefore, aimed to investigate the interaction between foods and beverages and genetic predisposition to these three metabolic traits on the risk of developing T2D, by using three GRS.

## *6.3 Methods*

The EPIC-InterAct cohort, genotyping and imputation, dietary assessment and statistical analyses have already been described in the Methods Chapter (Chapter

3). In addition to these general methods, below I describe some specific methods for this piece of work.

*Cohort*

The current analyses were based on 21,900 adults with available genome-wide genotyping and dietary data (with 9,742 T2D cases including those in the subcohort and 12,158 non-cases).

*Genotyping and genetic risk score*

The same GRS as those used in the interaction analyses between GRS and macronutrient intake on incident T2D (Chapter 5) were used here.

*Self-reported dietary intake*

The habitual consumption of 15 foods and beverages that were previously investigated for their association with T2D in published systematic reviews and meta-analyses were examined (Table 6-1). Dietary data on habitual consumption over the previous year (g/day) was derived from self or interviewer-administered country-specific food frequency questionnaires or dietary histories taken at baseline. Validity of the instruments to capture habitual diet was assessed in each participating cohort.[159,226] All 15 foods and beverages were available in EPIC-InterAct, except for legume intake which was not available in Denmark.

| Table 5-1: Definition of foods and beverages included in the analyses: EPIC-InterAct study | |
|---|---|
| **Foods and beverages included (reference to published meta-analysis)** | **Definition and examples** |
| Fruits[44] | apples, oranges, grape, stone fruits, mixed fruits (dried fruits, fruit salad) |
| Green leafy vegetables[44] | lettuce, spinach, Swiss chard leaf |
| Root vegetables[44] | carrot, celeriac, radish, parsnip, beetroot. Exclude potato |
| Wholegrain breads and cereals[45] | bread and crispbreads (non-white), grains (e.g. couscous, semolina, polenta, pearl barley) |
| Rice[54] | white and brown rice |
| Legumes[46] | red kidney beans, chickpeas, lentils |
| Nuts and seeds[46] | nuts and seeds, tree nuts, peanuts, chestnuts |
| Fermented dairy[47,48] | yoghurt and fermented milk (e.g. kefir), cheese (e.g. ricotta, cheddar) |
| Unprocessed red meat[55] | beef, lamb, pork, goat |
| Processed red meat[55] | bacon, ham, cold meat |
| Fish[51] | fish including oily and white fish but not including crustaceans/fish products |
| Egg and egg products[52] | egg, pickled egg, egg powder |
| Sugar Sweetened Beverages (SSB)[53] | carbonated, soft, isotonic drinks, diluted syrups (e.g. cola, lemonade, sweetened or sugar reduced). Based on the proxy: total soft drinks minus artificially sweetened beverages |
| Coffee[49] | caffeinated, decaffeinated and partially caffeinated |
| Tea[50] | black and green teas |

*Assessment of covariates*

The method is the same as that described in the interaction analyses between GRS and macronutrient intake on incident T2D, in Chapter 5.

*Statistical analyses*

Variables with < 30% missing data (BMI, physical activity, education level and smoking status: Table 5-1) were imputed using multiple imputation by chained equations in *Stata*, as previously described,[210] to reduce bias from loss of data. After confirming no obvious between-imputation variation across 20 multiple-imputations, a single imputation was used for analyses because of computational

efficiency (Figure 6-1). Imputation was based on 21,900 individuals. Potential outliers were Winsorised at the 99th percentile.

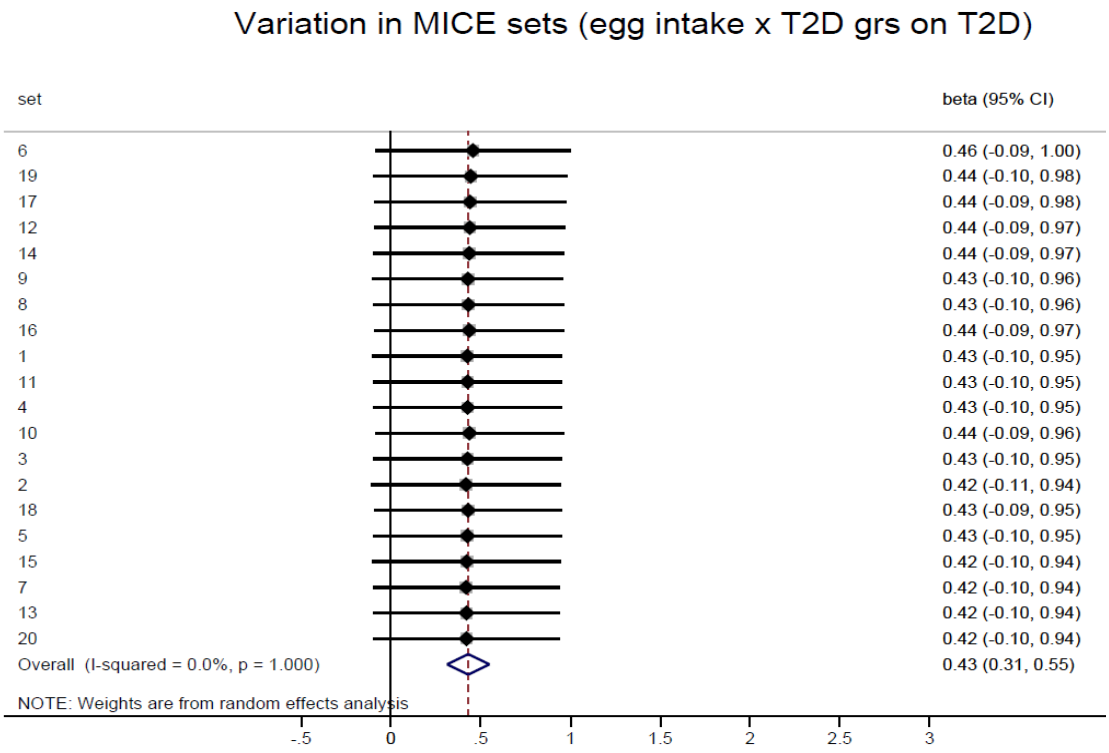



**Figure 5-1: Comparison between 20 imputed datasets for the interaction between egg and egg product intake with T2D genetic risk score and incident T2D**
Modelling same as that for model 4 within Table 6-4. Here, analysis performed by country and pooled using random effects meta-analysis (no chip-specific analysis was undertaken).
Total variation= within + between study variation is very small.
Within variance: 0.072
Between variance: 8.98E-5
**Conclusion: total variation= within + between study variation is very small. Dataset number 15 was therefore chosen, at random, for all analyses presented in this paper.**
Abbreviation: MICE: multiple imputation using chained equations, T2D: type 2 diabetes, GRS: genetic risk score

### *Main effect analyses*

The associations between each of the 15 foods and beverages and incident T2D was analysed by treating dietary exposures as continuous variables (per portion of intake/day as specified in Table 6-2). Average portion sizes were used to reflect a realistic amount of consumption and were derived from consultation of the Food Standards Agency 'Food Portion Sizes'[227] and published meta-analyses of the respective foods or beverages and T2D. Crude and multivariable-adjusted Prentice-weighted Cox regression models were constructed within country. Multivariable-adjusted models were constructed, with model 1 adjusting for age (years) as the underlying timescale, sex, centre (nominal categorical), total energy

intake (kcal/day), physical activity (inactive, moderately inactive, moderately active, active), education (none, primary school, technical/professional, secondary school, longer education (including university), smoking (never, former, current smoker), sex-specific alcohol intake (none, light drinking: 0.1-6g/d of pure alcohol, moderate drinking: men 6.1-24g/d and women 6.1-12g/d, heavy drinking: men>24g/d and women>12g/d), and mutual adjustment for all other foods except legumes (due to data availability) and model 2 additionally adjusting for BMI (kg/m$^2$). Country-specific hazard ratio (HR) for each dietary exposure and incident T2D was combined using random-effects meta-analysis.

Whereas for the main associations between GRS and T2D, the GRS were treated as both continuous (per SD difference) and dichotomised exposures (being high and low GRS estimated based on the median estimates from those in the subcohort). Prentice-weighted Cox regression models were constructed within country and by genotyping chip. Genotype chip-specific and then country-specific estimates were combined using random-effects meta-analysis. Models included adjustment for age (underlying time scale), sex, centre, the first 5 principal components for genetic population stratification and BMI.

*Interaction analyses*

All interaction findings and P values for interaction reported in the text and tables (Tables 6-4 to Table 6-6) were estimated by treating dietary (per 100g/day difference) and GRS exposures as continuous variables (per SD difference). Interactions were analysed using two types of statistical interaction approaches by introducing a product term between the GRS and dietary exposure. First, a multiplicative interaction was considered present if the HR for food or beverages and incident T2D differed between GRS groups (i.e. $RR_{GRS \times diet}/RR_{GRS} \times RR_{diet} < 1$ or >1). Second, additive interactions were tested by evaluating relative excess risk due to interaction (RERI), which can be useful in public health decision making [135]. The RERI examines whether an absolute incidence rate differs between GRS groups (i.e. $RR_{GRS \times diet} - RR_{GRS} - RR_{diet} + 1$ is < or >0).[135] Regression models were constructed in the same way as in the analysis of the association between GRS and incident T2D, and the list of covariates included in the models were the same as for the main associations between dietary exposures and T2D (described

above), with addition of the first 5 principal components for population stratification. Between-country heterogeneity was quantified by the $I^2$ value and P for heterogeneity was derived from the Cochran-Q test. The modelling for the multiplicative interaction analyses resulted in four levels, with sequential addition of more covariates (Table 6-4). Whereas, in additive interaction analyses, only most adjusted multivariable model has been presented (Table 6-5). BMI was a covariate in the interaction analysis for IR and T2D GRS only. To account for multiple testing for 15 dietary exposures and 3 GRS, the p-value threshold was set as 0.001 (0.05/45) for each of the multiplicative and additive interaction analyses. To visualise potential interactions, we separately estimated the prospective association between dietary exposures and T2D, stratified by those possessing low or high genetic risk for each of the three GRS (Figure 6-2).

Given that GRS may mask interactions between diet and individual SNPs, secondary interaction analysis was conducted for each of 195 non-overlapping SNPs within all 3 GRS and 15 foods and beverages (for which p<1.71E-5 was designated as significant). *Stata* version 14 (StataCorp LP, Texas, USA) was used for analysis.

## *6.4 Results*

The demographic data and association between the three GRS and incident T2D are the same as that described in Chapter 5 (Table 6-2). In the adjusted analyses, there was a positive association between habitual processed red meat intake (HR per portion/d: 1.22; 95%CI 1.09,1.37) and habitual SSB intake (HR per can/d: 1.21;95%CI 1.02,1.43) on incident T2D, whereas habitual coffee (HR per cup/d: 0.92;95%CI 0.89,0.95) and habitual tea intake (HR per cup/d: 0.95;95%CI 0.90,0.999) were inversely associated with T2D (Table 6-2). There was an overall null association with other dietary factors. There was little correlation between the GRS and dietary factors (Table 6-3).

**Table 5-2: Association between food and beverage intake or genetic risk score with incident T2D: EPIC-InterAct study**

| | **Mean (SD)** | **Mean (SD)** | **HR (95%CI) per portion/d** | |
|---|---|---|---|---|
| | subcohort non-cases | total incident cases | **Model 1** | **Model 2** |
| **Median years of follow up (y)** | 12.3 | 6.8 | | |
| **Age at baseline (y)** | 52.3 (9.3) | 55.7 (7.6) | | |
| **Sex (%male)** | 37.9 | 49.9 | | |
| **Foods or beverages (portion size)** | **Median (IQR)** | **Median (IQR)** | | |
| Fruit (100 g/d) | 187.8(102,307) | 178.6(95.9,300.8) | 1.02(0.98,1.06) | 0.98(0.94,1.03) |
| Green leafy vegetables (90 g/d) | 12.0(2.6,33.4) | 8.6(1.4,30.4) | 1.19(0.96,1.47) | 1.11(0.99,1.25) |
| Root vegetables (80 g/d) | 11.9(4.4,28.8) | 10.6(3.9,27.3) | 0.83(0.71,0.96) | 0.90(0.77,1.05) |
| Wholegrain breads/cereals (40 g/d) | 42.3(2.9,102.5) | 44.3(2.5,102.5) | 0.99(0.96,1.02) | 1.00(0.96,1.04) |
| Rice (100 g/d) | 15.1(5.3,28.7) | 14.3(3.5,26.0) | 0.97(0.80,1.18) | 0.75(0.48,1.18) |
| Fermented dairy (125 ml/d) | 73.7(34.5,140.7) | 66.6(28.6,131.7) | 0.97(0.92,1.01) | 0.95(0.90,1.02) |
| Nuts and seeds (30 g/d) | 0.7(0,3.0) | 0.3(0,1.7) | 1.00(0.74,1.33) | 1.05(0.85,1.29) |
| Legumes (35 g/d)* | 6.2(0.7,23.1) | 6.2(0.5,23.1) | 1.14(0.96,1.35) | 1.11(0.90,1.37) |
| Red meat (144 g/d) | 37.7(18.3,65.6) | 43.3(22.5,71.2) | 1.78(1.36,2.34) | 1.20(0.99,1.45) |
| Processed red meat (75 g/d) | 28.5(14.9,49.5) | 32.7(17.6,56.1) | 1.46(1.30,1.65) | 1.22(1.09,1.37) |
| Fish (100 g/d) | 5.9(0.6,14.4) | 6.6(0.6,14.9) | 1.26(0.91,1.75) | 1.09(0.82,1.46) |
| Egg and egg products (50 g/d) | 14.3(6.7,24.4) | 15.2(7.0,27.0) | 1.26(1.03,1.55) | 1.03(0.77,1.38) |
| SSB (336 ml/d) | 0(0,42.9) | 1.2(0,57.1) | 1.34(1.13,1.60) | 1.21(1.02,1.43) |
| Coffee (260 ml/d) | 298(98.9,580.2) | 300(95.7,600.0) | 0.93(0.90,0.97) | 0.92(0.89,0.95) |
| Tea (260 ml/d) | 5.1(0,200.0) | 2.5(0,150.0) | 0.91(0.87,0.94) | 0.95(0.90,0.999) |

| Unweighted genetic risk score | Sample size | Model 1 | Model 2 |
|---|---|---|---|
| GRS for T2D | | | |
| per SD (4.3 risk alleles) | | 1.42(1.32,1.52) | 1.49(1.37,1.63) |
| High (≥52 risk alleles) | 12033 | 1.75(1.55,1.98) | 1.85(1.58,2.16) |
| Low (<52) | 9867 | 1 [Reference] | 1 [Reference] |
| GRS for IR | | | |
| per SD (4.5 risk alleles) | | 1.11(1.08,1.14) | 1.14(1.09,1.20) |
| High (≥55) | 11303 | 1.18(1.12,1.25) | 1.23(1.12,1.35) |
| Low (<55) | 10597 | 1 [Reference] | 1 [Reference] |
| GRS for BMI | | | |
| per SD (6.3 risk alleles) | | 1.07(1.04,1.10) | NA |
| High (≥91) | 11230 | 1.12(1.06,1.19) | |
| Low (<91) | 10670 | 1 [Reference] | |

**Hazard ratio (HR) for foods and beverages (per portion) and incident Type 2 Diabetes (T2D), adjusted for:** Model 1: age (=underlying time scale), sex, centre, total energy intake, physical activity, education, smoking status, sex-specific alcohol categories, all foods are mutually adjusted (except for legumes); Model 2: model 1 + BMI

**HR for the association between unweighted genetic risk score and incident T2D, adjusted for:** Model 1: age (=underlying time scale), sex, centre, first 5 principal components for population stratification; Model 2: model 1 + BMI. BMI GRS not adjusted for BMI.

Number of SNPs: T2D (48 SNPs from Morris et al., Nature Gen, 2012), BMI (97 as per Locke et al., Nature, 2015), IR (53 as per Lotta et al., Nature Gen, 2016)
Abbreviations: T2D- type 2 diabetes, BMI- body mass index, IR- insulin resistance, SD- standard deviation

*Sample size: all dietary exposures: 9,742 incident T2D cases and 21,900 participants in total. Except, for legumes, which had 7,946 incident T2D cases and 18,334 participants in total.

**Table 5-3: Correlations between exposures within the EPIC-InterAct study**

| | T2D GRS | IR GRS | BMI GRS | Fruit | Leafy Veg | Root Veg | Whole grain | Rice | Fermented Dairy | Legumes | Nuts and seeds | Red Meat | Proce-ssed Meat | Fish | Egg and egg products | SSB | Coffee | Tea |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Fruit** | 0.01 | 0.00 | 0.01 | | 0.34 | 0.11 | 0.16 | 0.13 | 0.08 | 0.22 | 0.01 | 0.04 | 0.16 | 0.16 | 0.06 | 0.10 | 0.27 | 0.07 |
| **Leafy Veg** | 0.01 | 0.01 | 0.00 | | | 0.09 | 0.34 | 0.20 | 0.05 | 0.39 | 0.05 | 0.02 | 0.08 | 0.09 | 0.18 | 0.23 | 0.40 | 0.29 |
| **Root Veg** | 0.00 | 0.00 | 0.01 | | | | 0.24 | 0.09 | 0.17 | 0.19 | 0.09 | 0.02 | 0.10 | 0.03 | 0.09 | 0.13 | 0.16 | 0.33 |
| **Wholegrain** | 0.01 | 0.00 | 0.00 | | | | | 0.23 | 0.18 | 0.37 | 0.19 | 0.09 | 0.08 | 0.15 | 0.03 | 0.12 | 0.39 | 0.36 |
| **Rice** | 0.01 | 0.00 | 0.02 | | | | | | 0.04 | 0.38 | 0.05 | 0.05 | 0.03 | 0.22 | 0.09 | 0.08 | 0.18 | 0.10 |
| **Fermented Dairy** | 0.00 | 0.01 | 0.00 | | | | | | | 0.20 | 0.07 | 0.07 | 0.02 | 0.02 | 0.05 | 0.06 | 0.12 | 0.10 |
| **Legumes** | 0.01 | 0.01 | 0.00 | | | | | | | | 0.06 | 0.10 | 0.06 | 0.25 | 0.21 | 0.20 | 0.34 | 0.11 |
| **Nuts and seeds** | 0.01 | 0.00 | 0.03 | | | | | | | | | 0.01 | 0.01 | 0.06 | 0.03 | 0.08 | 0.08 | 0.21 |
| **Red Meat** | 0.01 | 0.00 | 0.01 | | | | | | | | | | 0.15 | 0.18 | 0.28 | 0.02 | 0.17 | 0.05 |
| **Processed Meat** | 0.01 | 0.01 | 0.00 | | | | | | | | | | | 0.09 | 0.16 | 0.07 | 0.11 | 0.12 |
| **Fish** | 0.02 | 0.01 | 0.01 | | | | | | | | | | | | 0.16 | 0.06 | 0.08 | 0.04 |
| **Egg and egg products** | 0.00 | 0.00 | 0.01 | | | | | | | | | | | | | 0.13 | 0.02 | 0.16 |
| **SSB** | 0.02 | 0.01 | 0.02 | | | | | | | | | | | | | | 0.17 | 0.22 |
| **Coffee** | 0.01 | 0.00 | 0.02 | | | | | | | | | | | | | | | 0.14 |
| **Tea** | 0.01 | 0.00 | 0.01 | | | | | | | | | | | | | | | |


Spearman's rho without direction of correlation, all variables treated as continuous variables.
Abbreviations: Veg- vegetables, SSB- sugar sweetened beverage
Correlations calculated based on subcohort population only (N=12,749)
**Effective number of independent tests**:[203] 14.596 (variance of the observed eigenvalues: 0.433). Therefore, 15 tests conducted per GRS
http://gump.qimr.edu.au/general/daleN/matSpD/
This estimates the number of independent tests, given the correlation between exposures (provided in the table above).

### Interactions between three GRS and food and beverage intake on incident T2D

Under the multiplicative interaction scale, two interactions showed a nominally significant interaction ($p_{interaction}<0.05$): eggs and egg product intake and T2D GRS ($p_{interaction}=0.039$); and SSB and BMI GRS ($p_{interaction}=0.012$) on T2D risk (Figure 6-2 and Table 6-4). The interaction between egg and egg products and the T2D GRS showed high heterogeneity between countries ($I^2$:80%). However, no interactions passed multiple testing corrections. Results also did not change materially after adjustment for different sets of confounders using multivariable adjusted models (Table 6-4). Similarly this was the case for when T2D and BMI GRS were weighted by the beta coefficient of each SNPs effect on their respective trait from published GWAS meta-analyses. On the additive interaction scale, an interaction was also noted between SSB and BMI GRS on T2D risk ($p_{interaction}=0.02$), however no interactions were significant after accounting for multiple testing (Table 6-5).

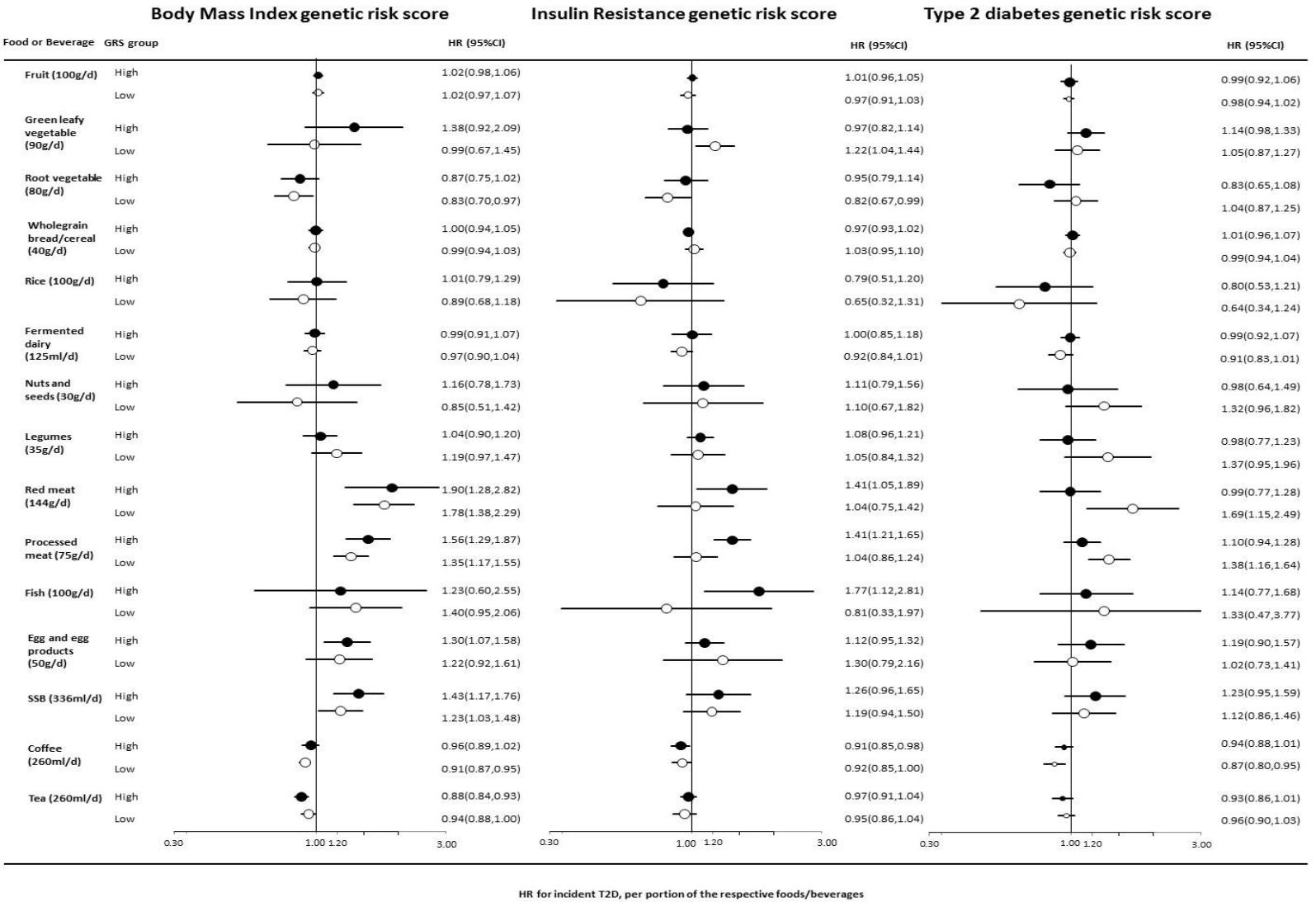

Body Mass Index genetic risk score
Insulin Resistance genetic risk score
Type 2 diabetes genetic risk score
Food or Beverage
GRS group
HR (95%CI)
HR (95%CI)
HR (95%CI)
Fruit (100g/d)
High 1.02(0.98,1.06) 1.01(0.96,1.05) 0.99(0.92,1.06)
Low 1.02(0.97,1.07) 0.97(0.91,1.03) 0.98(0.94,1.02)
Green leafy vegetable (90g/d)
High 1.38(0.92,2.09) 0.97(0.82,1.14) 1.14(0.98,1.33)
Low 0.99(0.67,1.45) 1.22(1.04,1.44) 1.05(0.87,1.27)
Root vegetable (80g/d)
High 0.87(0.75,1.02) 0.95(0.79,1.14) 0.83(0.65,1.08)
Low 0.83(0.70,0.97) 0.82(0.67,0.99) 1.04(0.87,1.25)
Wholegrain bread/cereal (40g/d)
High 1.00(0.94,1.05) 0.97(0.93,1.02) 1.01(0.96,1.07)
Low 0.99(0.94,1.03) 1.03(0.95,1.10) 0.99(0.94,1.04)
Rice (100g/d)
High 1.01(0.79,1.29) 0.79(0.51,1.20) 0.80(0.53,1.21)
Low 0.89(0.68,1.18) 0.65(0.32,1.31) 0.64(0.34,1.24)
Fermented dairy (125ml/d)
High 0.99(0.91,1.07) 1.00(0.85,1.18) 0.99(0.92,1.07)
Low 0.97(0.90,1.04) 0.92(0.84,1.01) 0.91(0.83,1.01)
Nuts and seeds (30g/d)
High 1.16(0.78,1.73) 1.11(0.79,1.56) 0.98(0.64,1.49)
Low 0.85(0.51,1.42) 1.10(0.67,1.82) 1.32(0.96,1.82)
Legumes (35g/d)
High 1.04(0.90,1.20) 1.08(0.96,1.21) 0.98(0.77,1.23)
Low 1.19(0.97,1.47) 1.05(0.84,1.32) 1.37(0.95,1.96)
Red meat (144g/d)
High 1.90(1.28,2.82) 1.41(1.05,1.89) 0.99(0.77,1.28)
Low 1.78(1.38,2.29) 1.04(0.75,1.42) 1.69(1.15,2.49)
Processed meat (75g/d)
High 1.56(1.29,1.87) 1.41(1.21,1.65) 1.10(0.94,1.28)
Low 1.35(1.17,1.55) 1.04(0.86,1.24) 1.38(1.16,1.64)
Fish (100g/d)
High 1.23(0.60,2.55) 1.77(1.12,2.81) 1.14(0.77,1.68)
Low 1.40(0.95,2.06) 0.81(0.33,1.97) 1.33(0.47,3.77)
Egg and egg products (50g/d)
High 1.30(1.07,1.58) 1.12(0.95,1.32) 1.19(0.90,1.57)
Low 1.22(0.92,1.61) 1.30(0.79,2.16) 1.02(0.73,1.41)
SSB (336ml/d)
High 1.43(1.17,1.76) 1.26(0.96,1.65) 1.23(0.95,1.59)
Low 1.23(1.03,1.48) 1.19(0.94,1.50) 1.12(0.86,1.46)
Coffee (260ml/d)
High 0.96(0.89,1.02) 0.91(0.85,0.98) 0.94(0.88,1.01)
Low 0.91(0.87,0.95) 0.92(0.85,1.00) 0.87(0.80,0.95)
Tea (260ml/d)
High 0.88(0.84,0.93) 0.97(0.91,1.04) 0.93(0.86,1.01)
Low 0.94(0.88,1.00) 0.95(0.86,1.04) 0.96(0.90,1.03)
0.30
1.00 1.20
3.00
HR for incident T2D, per portion of the respective foods/beverages

**Figure 5-2: Prospective associations of food or beverage intake (per portion) with incidence of type 2 diabetes, stratified by genetic risk for the respective metabolic traits: EPIC-InterAct study**

Adjusted for age (=underlying time scale), sex, centre, total energy intake, first five principal components for population stratification, physical activity, education, smoking, sex-specific alcohol categories, body-mass index and mutually adjusted for all other foods (except legumes). See Model 4 of Table 6-4 and methods for more information. Median number of risk alleles for genetic risk score (GRS) group: T2D GRS (low<52 risk alleles); IR GRS (low<55 risk alleles); BMI GRS (low<91 risk alleles). Between country heterogeneity ($I^2$) ranged from 0 to 80% (Table 6-4). There are no statistically significant interactions (p value threshold<0.001 based on 0.05/45 tests). P values for interactions are provided in Table 6-4 and were estimated using continuous exposures: per 100g/d of dietary exposure and per SD difference in GRS. The lowest p value for interaction for each respective GRS: between T2D GRS and egg and egg product intake ($p_{interaction}$=0.039), IR GRS and legume intake ($p_{interaction}$=0.056) and BMI GRS and SSB intake ($p_{interaction}$=0.012).

Abbreviations: SSB: sugar sweetened beverage; GRS: genetic risk score; HR: hazard ratio. Black dots: high genetic risk group, white dots: low genetic risk group.

**Table 5-4: Multiplicative interaction between three metabolic genetic risk scores and food or beverage intakes on incident T2D: EPIC-InterAct study (Unweighted GRS)**

| Food or beverage (100 g/d) | Model | GRS for body mass index (per 6.3 risk alleles)[%] | | | GRS for insulin resistance (per 4.5 risk alleles) | | | GRS for type 2 diabetes (per 4.3 risk alleles) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Beta (95% CI) | *P* | $I^2$ (%) | Beta (95% CI) | *P* | $I^2$ (%) | Beta (95% CI) | *P* | $I^2$ (%) |
| Fruit | *model 1* | 0.013(-0.006,0.032) | 0.168 | 0.0 | -0.001(-0.025,0.023) | 0.951 | 25.4 | 0.006(-0.026,0.039) | 0.704 | 56.2 |
| | *model 2* | 0.013(-0.009,0.036) | 0.242 | 16.0 | -0.001(-0.026,0.023) | 0.926 | 25.6 | 0.003(-0.03,0.035) | 0.875 | 55.3 |
| | *model 3* | 0.01(-0.012,0.033) | 0.355 | 10.3 | 0(-0.025,0.024) | 0.972 | 19.8 | 0.003(-0.032,0.037) | 0.881 | 55.6 |
| | *model 4* | | | | -0.004(-0.037,0.03) | 0.830 | 41.0 | 0.01(-0.035,0.055) | 0.663 | 67.1 |
| Green leafy vegetable | *model 1* | 0.076(-0.024,0.175) | 0.137 | 0.0 | -0.074(-0.16,0.011) | 0.088 | 0.0 | 0.098(-0.182,0.379) | 0.493 | 53.1 |
| | *model 2* | 0.082(-0.02,0.185) | 0.115 | 0.0 | -0.072(-0.16,0.017) | 0.113 | 0.0 | 0.044(-0.244,0.332) | 0.767 | 52.8 |
| | *model 3* | 0.071(-0.056,0.198) | 0.274 | 2.4 | -0.049(-0.142,0.044) | 0.298 | 0.0 | 0.089(-0.177,0.354) | 0.513 | 39.8 |
| | *model 4* | | | | -0.075(-0.173,0.023) | 0.135 | 0.0 | 0.351(-0.016,0.718) | 0.061 | 56.3 |
| Root vegetable | *model 1* | -0.004(-0.12,0.111) | 0.940 | 0.0 | 0.085(-0.038,0.207) | 0.176 | 0.0 | 0.001(-0.118,0.119) | 0.992 | 0.0 |
| | *model 2* | 0.013(-0.105,0.131) | 0.833 | 0.0 | 0.065(-0.072,0.202) | 0.351 | 6.2 | -0.01(-0.129,0.109) | 0.870 | 0.0 |
| | *model 3* | 0.014(-0.108,0.137) | 0.819 | 0.0 | 0.087(-0.044,0.218) | 0.194 | 0.0 | -0.037(-0.16,0.086) | 0.553 | 0.0 |
| | *model 4* | | | | 0.061(-0.092,0.214) | 0.435 | 2.4 | -0.12(-0.252,0.013) | 0.077 | 0.0 |
| Whole grain breads and cereals | *model 1* | 0.017(-0.052,0.087) | 0.625 | 29.4 | 0.001(-0.08,0.083) | 0.980 | 48.6 | -0.016(-0.069,0.037) | 0.555 | 0.0 |
| | *model 2* | 0.019(-0.047,0.085) | 0.566 | 20.7 | -0.014(-0.096,0.067) | 0.733 | 45.8 | -0.02(-0.075,0.034) | 0.459 | 0.0 |
| | *model 3* | 0.002(-0.068,0.073) | 0.951 | 21.9 | -0.017(-0.108,0.074) | 0.711 | 50.4 | -0.018(-0.075,0.039) | 0.539 | 0.0 |
| | *model 4* | | | | -0.046(-0.148,0.055) | 0.367 | 49.6 | 0.003(-0.062,0.067) | 0.935 | 0.0 |
| Rice | *model 1* | 0.051(-0.162,0.264) | 0.638 | 38.9 | -0.065(-0.248,0.119) | 0.490 | 29.8 | 0.062(-0.088,0.211) | 0.417 | 0.0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *model 2* | 0.017(-0.186,0.219) | 0.871 | 30.3 | -0.086(-0.253,0.082) | 0.316 | 16.5 | 0.05(-0.102,0.203) | 0.517 | 0.0 |
| | *model 3* | 0.061(-0.164,0.286) | 0.595 | 35.9 | -0.072(-0.283,0.139) | 0.504 | 37.7 | 0.063(-0.096,0.221) | 0.441 | 0.0 |
| | *model 4* | | | | -0.003(-0.254,0.248) | 0.983 | 41.8 | 0.17(-0.051,0.392) | 0.132 | 20.4 |
| Legumes† | *model 1* | -0.065(-0.219,0.09) | 0.413 | 0.0 | 0.034(-0.128,0.196) | 0.678 | 0.0 | 0.011(-0.285,0.307) | 0.943 | 31.4 |
| | *model 2* | -0.084(-0.243,0.076) | 0.302 | 0.0 | 0.051(-0.118,0.219) | 0.557 | 0.0 | -0.03(-0.35,0.29) | 0.853 | 35.4 |
| | *model 3* | -0.064(-0.232,0.104) | 0.457 | 0.0 | 0.062(-0.112,0.236) | 0.488 | 0.0 | -0.059(-0.487,0.369) | 0.786 | 54.3 |
| | *model 4* | | | | 0.186(-0.004,0.376) | 0.056 | 0.0 | -0.038(-0.67,0.593) | 0.905 | 70.2 |
| Nuts and seeds (per 30 g/d) | *model 1* | 0.019(-0.152,0.19) | 0.830 | 6.7 | 0.08(-0.081,0.24) | 0.332 | 0.0 | 0.008(-0.15,0.167) | 0.917 | 0.0 |
| | *model 2* | -0.039(-0.233,0.154) | 0.692 | 14.0 | 0.109(-0.058,0.276) | 0.202 | 0.0 | 0.031(-0.182,0.245) | 0.774 | 22.7 |
| | *model 3* | -0.086(-0.385,0.212) | 0.571 | 53.2 | 0.148(-0.026,0.322) | 0.096 | 0.0 | 0.051(-0.176,0.278) | 0.662 | 22.6 |
| | *model 4* | | | | 0.046(-0.156,0.248) | 0.655 | 0.0 | -0.073(-0.456,0.309) | 0.707 | 58.8 |
| Fermented dairy | *model 1* | 0.002(-0.033,0.036) | 0.932 | 0.6 | 0.028(-0.041,0.098) | 0.427 | 67.3 | 0.04(0.005,0.076) | 0.024 | 0.0 |
| | *model 2* | 0.009(-0.026,0.044) | 0.612 | 0.0 | 0.029(-0.045,0.103) | 0.441 | 69.7 | 0.042(0.006,0.078) | 0.023 | 0.8 |
| | *model 3* | 0.011(-0.025,0.047) | 0.543 | 0.0 | 0.027(-0.053,0.107) | 0.514 | 71.7 | 0.037(-0.008,0.082) | 0.108 | 19.7 |
| | *model 4* | | | | 0.002(-0.064,0.068) | 0.956 | 45.9 | 0.028(-0.032,0.087) | 0.365 | 33.0 |
| Red meat | *model 1* | -0.016(-0.103,0.072) | 0.725 | 0.0 | 0.013(-0.073,0.099) | 0.769 | 0.0 | -0.028(-0.118,0.061) | 0.536 | 0.0 |
| | *model 2* | 0.002(-0.088,0.093) | 0.959 | 0.0 | 0.013(-0.076,0.103) | 0.767 | 0.0 | -0.043(-0.135,0.049) | 0.359 | 0.0 |
| | *model 3* | 0.018(-0.079,0.114) | 0.717 | 0.0 | 0.015(-0.078,0.108) | 0.753 | 0.0 | -0.043(-0.139,0.053) | 0.377 | 0.0 |
| | *model 4* | | | | 0.024(-0.081,0.128) | 0.660 | 0.0 | -0.082(-0.245,0.081) | 0.323 | 45.3 |
| Processed red meat | *model 1* | 0.078(-0.084,0.24) | 0.344 | 55.5 | 0.006(-0.091,0.103) | 0.903 | 0.0 | -0.062(-0.232,0.107) | 0.470 | 58.1 |
| | *model 2* | 0.089(-0.074,0.252) | 0.286 | 53.2 | 0.034(-0.067,0.136) | 0.505 | 0.0 | -0.064(-0.22,0.092) | 0.420 | 47.2 |
| | *model 3* | 0.094(-0.076,0.265) | 0.277 | 51.6 | 0.032(-0.074,0.138) | 0.551 | 0.0 | -0.08(-0.262,0.102) | 0.390 | 56.3 |
| | *model 4* | | | | 0.023(-0.098,0.144) | 0.709 | 0.0 | -0.106(-0.319,0.107) | 0.330 | 54.7 |
| Fish | *model 1* | -0.022(-0.27,0.226) | 0.863 | 0.0 | 0.2(-0.116,0.517) | 0.214 | 21.9 | 0.235(-0.231,0.702) | 0.323 | 55.3 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *model 2* | -0.033(-0.288,0.221) | 0.798 | 0.0 | 0.176(-0.214,0.566) | 0.376 | 38.9 | 0.23(-0.291,0.75) | 0.387 | 61.5 |
| | *model 3* | -0.002(-0.269,0.264) | 0.988 | 0.0 | 0.223(-0.166,0.613) | 0.262 | 35.4 | 0.303(-0.251,0.857) | 0.283 | 63.1 |
| | *model 4* | | | | 0.407(-0.224,1.038) | 0.207 | 65.5 | 0.26(-0.542,1.062) | 0.525 | 76.6 |
| Egg and egg products | *model 1* | -0.094(-0.27,0.082) | 0.293 | 0.0 | -0.067(-0.36,0.226) | 0.655 | 55.5 | 0.127(-0.099,0.354) | 0.271 | 16.9 |
| | *model 2* | -0.053(-0.264,0.157) | 0.619 | 15.4 | -0.072(-0.395,0.252) | 0.663 | 59.9 | 0.126(-0.09,0.342) | 0.252 | 7.2 |
| | *model 3* | -0.025(-0.248,0.199) | 0.829 | 15.2 | -0.089(-0.415,0.236) | 0.590 | 51.9 | 0.202(-0.095,0.499) | 0.182 | 36.4 |
| | *model 4* | | | | -0.183(-0.58,0.213) | 0.365 | 56.8 | 0.638(0.032,1.243) | 0.039 | 79.8* |
| SSB | *model 1* | 0.027(0,0.053) | 0.047 | 0.0 | -0.006(-0.033,0.021) | 0.650 | 0.0 | -0.005(-0.035,0.025) | 0.729 | 13.2 |
| | *model 2* | 0.027(0,0.055) | 0.054 | 0.0 | -0.006(-0.034,0.022) | 0.694 | 0.0 | 0(-0.042,0.041) | 0.982 | 42.6 |
| | *model 3* | 0.037(0.008,0.066) | 0.012 | 0.0* | -0.001(-0.031,0.028) | 0.931 | 0.0 | -0.001(-0.037,0.036) | 0.969 | 26.9 |
| | *model 4* | | | | 0.007(-0.03,0.044) | 0.711 | 13.4 | 0.001(-0.053,0.056) | 0.963 | 54.0 |
| Coffee | *model 1* | 0.008(-0.007,0.023) | 0.314 | 41.8 | 0.001(-0.017,0.018) | 0.946 | 58.1 | 0.017(-0.004,0.038) | 0.111 | 64.9 |
| | *model 2* | 0.008(-0.011,0.027) | 0.400 | 55.9 | -0.001(-0.016,0.015) | 0.938 | 46.0 | 0.021(-0.001,0.043) | 0.065 | 65.8 |
| | *model 3* | 0.009(-0.01,0.029) | 0.347 | 54.5 | -0.002(-0.019,0.015) | 0.802 | 48.3 | 0.017(-0.007,0.041) | 0.161 | 67.4 |
| | *model 4* | | | | -0.001(-0.023,0.02) | 0.891 | 55.1 | 0.027(-0.006,0.059) | 0.109 | 76.7 |
| Tea | *model 1* | -0.006(-0.018,0.006) | 0.305 | 0.0 | -0.003(-0.016,0.01) | 0.622 | 0.0 | 0.004(-0.01,0.017) | 0.610 | 0.0 |
| | *model 2* | -0.006(-0.019,0.006) | 0.340 | 0.0 | -0.003(-0.016,0.01) | 0.634 | 0.0 | 0.003(-0.011,0.017) | 0.674 | 0.0 |
| | *model 3* | -0.009(-0.022,0.005) | 0.214 | 0.6 | -0.005(-0.019,0.009) | 0.500 | 0.0 | 0.003(-0.012,0.018) | 0.685 | 0.0 |
| | *model 4* | | | | -0.012(-0.036,0.012) | 0.333 | 24.9 | 0(-0.024,0.025) | 0.971 | 29.2 |

Beta coefficient for the interaction between the respective GRS and foods or beverages on incident T2D, adjusted for :
Model 1: age (=underlying time scale), sex, centre, total energy intake, first 5 principal components for population stratification; Model 2: model 1 + physical activity, education, smoking, sex-specific alcohol categories; Model 3: model 2+ mutually adjusted for all other foods except for legumes; Model 4: model 3+ BMI
**α for significant interaction: 0.001 (0.05/45 tests)**
*$p_{interaction}<0.05$
% Interaction analyses using the BMI GRS does not include adjustment for BMI because this would remove the variance explained by the exposure
† N=7946 incident T2D cases/18334 total participants; Note: model 4 for nut: does not adjust for alcohol intake due to convergence; model 4 interaction between IR GRS and green leafy vegetables, rice or dairy does not adjust for alcohol due to convergence

**Table 5-5: Additive interactions between three genetic risk scores and foods and beverages on incident T2D: EPIC-InterAct study**

| Food or beverage (100 g/d) | GRS for body mass index (per 6.3 risk alleles)[%] | | | GRS for insulin resistance (per 4.5 risk alleles) | | | GRS for type 2 diabetes (per 4.3 risk alleles) | | |
|---|---|---|---|---|---|---|---|---|---|
| | **RERI (95% CI)** | ***P*** | **$I^2$ (%)** | **RERI (95% CI)** | ***P*** | **$I^2$ (%)** | **RERI (95% CI)** | ***P*** | **$I^2$ (%)** |
| Fruit | 0.011(-0.01,0.032) | 0.300 | 0.0 | -0.007(-0.04,0.026) | 0.670 | 33.9 | -0.003(-0.056,0.05) | 0.910 | 51.5 |
| Green leafy vegetable | 0.058(-0.068,0.183) | 0.370 | 0.0 | -0.088(-0.197,0.021) | 0.120 | 0.0 | 0.246(0.059,0.432) | 0.010 | 0.0* |
| Root vegetable | -0.015(-0.116,0.087) | 0.770 | 0.0 | -0.026(-0.163,0.111) | 0.710 | 6.8 | -0.252(-0.478,-0.025) | 0.030 | 34.2* |
| Whole grain breads and cereals | 0(-0.058,0.058) | 1.000 | 0.0 | -0.064(-0.134,0.006) | 0.070 | 0.0 | -0.021(-0.119,0.077) | 0.670 | 2.2 |
| Rice | 0.019(-0.177,0.214) | 0.850 | 26.3 | -0.02(-0.137,0.097) | 0.740 | 0.0 | -0.136(-0.418,0.146) | 0.340 | 48.0 |
| Legumes† | -0.089(-0.265,0.087) | 0.320 | 0.0 | 0.116(-0.081,0.313) | 0.250 | 0.0 | -0.094(-0.369,0.18) | 0.500 | 0.0 |
| Nuts and seeds (per 30 g/d) | -0.084(-0.262,0.093) | 0.350 | 11.0 | -0.036(-0.255,0.182) | 0.750 | 0.0 | -0.327(-0.637,-0.016) | 0.040 | 23.0* |
| Fermented dairy | 0(-0.037,0.036) | 0.990 | 0.0 | -0.008(-0.054,0.039) | 0.750 | 1.7 | -0.018(-0.136,0.1) | 0.770 | 69.6 |
| Red meat | 0.036(-0.098,0.17) | 0.590 | 0.0 | 0.043(-0.092,0.178) | 0.530 | 0.0 | -0.079(-0.257,0.099) | 0.390 | 9.1 |
| Processed red meat | 0.101(-0.208,0.41) | 0.520 | 47.7 | 0.085(-0.098,0.268) | 0.360 | 0.0 | -0.046(-0.263,0.171) | 0.680 | 0.0 |
| Fish | -0.059(-0.372,0.254) | 0.710 | 0.0 | 0.073(-0.426,0.572) | 0.770 | 9.9 | -0.233(-0.969,0.503) | 0.530 | 48.4 |
| Egg and egg products | -0.173(-0.453,0.106) | 0.220 | 2.0 | -0.146(-0.351,0.06) | 0.160 | 1.4 | -0.223(-0.653,0.207) | 0.310 | 33.3 |
| SSB | 0.039(0.006,0.071) | 0.020 | 0.0* | -0.002(-0.04,0.037) | 0.930 | 0.0 | -0.028(-0.123,0.067) | 0.560 | 68.5 |
| Coffee | 0.006(-0.014,0.026) | 0.540 | 56.8 | -0.008(-0.031,0.014) | 0.460 | 50.6 | -0.01(-0.047,0.027) | 0.610 | 58.8 |
| Tea | -0.01(-0.024,0.005) | 0.190 | 0.0 | -0.01(-0.028,0.008) | 0.250 | 0.5 | 0.002(-0.032,0.036) | 0.930 | 29.5 |

RERI for the interaction between the respective GRS and foods or beverages on incident T2D, adjusted for age (=underlying time scale), sex, centre, total energy intake, first 5 principal components for population stratification, physical activity, education, smoking, sex-specific alcohol categories, BMI, mutually adjusted for all other foods except for legumes.

**α for significant interaction: 0.001 (0.05/45 tests)**

*$p_{interaction}$<0.05

% Interaction analyses using the BMI GRS does not include adjustment for BMI because this would remove the variance explained by the exposure

† N=7946 incident T2D cases/18334 total participants

In secondary analyses, from 2,925 tests, there were 121 pairs of individual SNP and food or beverage interactions that demonstrated a $p_{interaction}$ <0.05 (Appendix H). None of these interactions remained statistically significant after accounting for multiple testing using either Bonferroni correction or the False Discovery Rate (Figure 6-3).[228] Moreover, results did not materially differ between our current analyses based on imputed data and a complete case analysis (9,403 cases and 11,745 non cases) (Table 6-6).

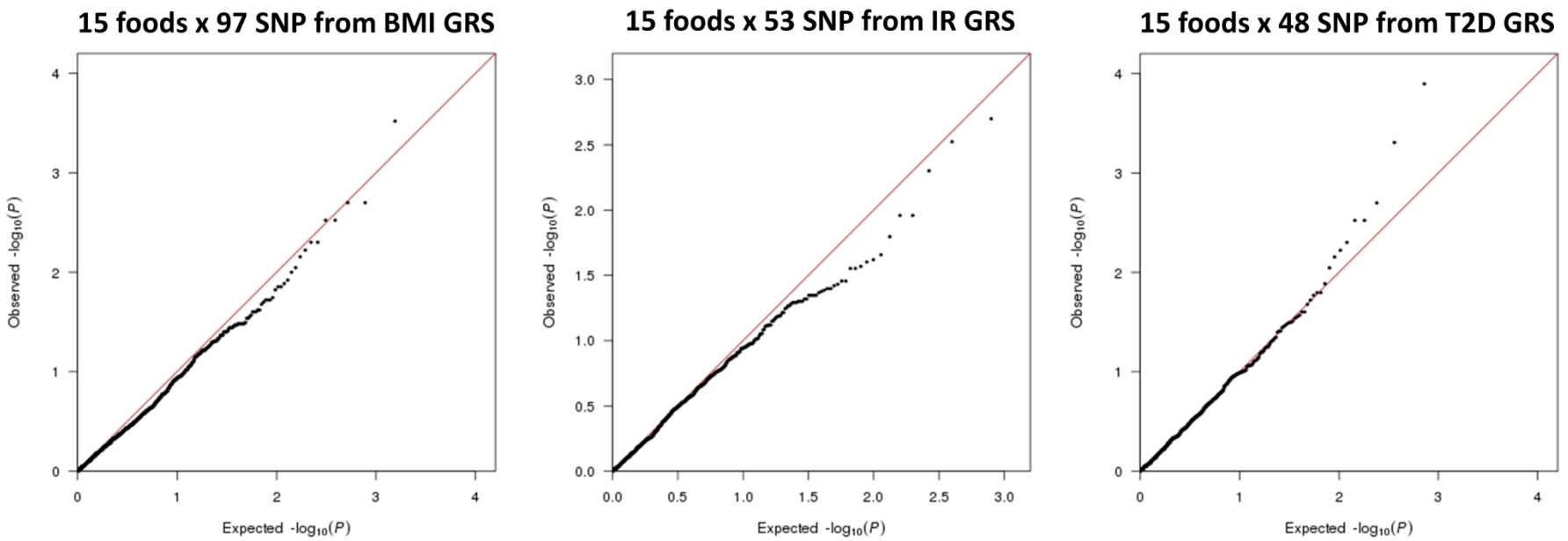



**Figure 5-3: QQ plot of p values for the interactions between individual SNPs for each of the genetic risk scores and food or beverage exposure on T2D: EPIC-InterAct Study**
All interactions are based on the most adjusted models (previously reported for Table 6-4). None of the individual SNP interactions were significant after accounting for multiple testing using Bonferroni correction p value threshold of p<1.71E-5 (based on 0.05/2,925 number of tests for 15 dietary exposures and non-overlapping SNPs across all 3 GRS). Please note that these analyses were conducted after excluding France (N=9,588 incident cases of T2D/21,418 total). Appendix H includes the results from individual SNP interactions with the respective dietary exposures that had a p value for interaction<0.05.

**Table 5-6: An example of interaction findings comparing multiple imputation and complete case analysis**

| Interaction | Multiple imputation analysis results (9742 cases, 12158 noncases) | | Complete case analysis results (9403 cases, 11745 noncases) | |
|---|---|---|---|---|
| | **Beta (se)** | ***P*** | **Beta (se)** | ***P*** |
| *Egg and egg products x unweighted T2D GRS* | 0.638 (0.309) | 0.039 | 0.631 (0.317) | 0.047 |
| SSB x unweighted BMI GRS | 0.037 (0.015) | 0.012 | 0.036 (0.015) | 0.017 |

Abbreviation: SSB: sugar sweetened beverage, BMI: body mass index, T2D: type 2 diabetes, GRS: genetic risk score, p: p value for interaction

**Conclusion: no substantial difference in results between analysis approaches**

## *6.5 Discussion*

In a large prospective study of T2D incidence across eight European countries included in EPIC-InterAct, the associations between foods or beverages on the risk of developing T2D does not significantly differ between those with low and high genetic risk for obesity, IR or T2D.

There are two main findings from these analyses. The main associations for the three GRS and T2D risk was consistent with previous literature.[5,28,30] They confirm known biological connections between adiposity and IR with the clinical manifestation of T2D. Similarly, the main associations for food and beverage intake and incident T2D were either consistent or not significant but directionally consistent with previously published EPIC-InterAct analyses and meta-analyses.[44–52,54,229–234]

We did not identify any statistically significant interactions between 15 foods or beverages and genes on the risk of T2D, on either the relative (multiplicative interaction analysis) or absolute (additive interaction analysis) scale. That is, in EPIC-InterAct, higher habitual consumption of a portion of processed red meat or a can of SSB raised the relative risk of T2D by 22% (HR: 1.22, 95%CI: 1.09,1.37) and 21% (HR: 1.21, 95%CI: 1.02,1.43) respectively, regardless of baseline genetic risk for obesity, IR or T2D (Figure 6-2). For example, even

among individuals with a lower genetic predisposition to T2D, those who habitually consumed a portion a day of processed red meat had 38% (HR 1.38; 95%CI: 1.16, 1.64) higher risk of developing future T2D compared to those that consumed none (Figure 6-2). This was in contrast to findings from a study among 2,533 US men[225], which was a smaller study (1,196 cases of T2D) than the current analyses with difficulties ascertaining directionality due to its case-control design. However, in another larger study (N=16,002 and replication N=21,421) examining BMI, a strong risk factor for T2D, the authors found the same adverse association between higher habitual consumption of fried food with higher BMI even among those with the lowest genetic predisposition for higher BMI.[235] Therefore, this suggests that current public health nutrition recommendations to reduce processed red meat and SSB intake apply to the whole population, irrespective of genetic predisposition. It was noted that the habitual consumption of a cup of coffee or tea was inversely associated with the relative risk of T2D by 8% (HR: 0.92, 95%CI: 0.89, 0.95) and 5% (HR: 0.95, 95%CI: 0.90, 0.999) respectively (Table 6-2), regardless of baseline genetic risk for obesity, IR or T2D (Figure 6-1). For example, those with a higher genetic predisposition for IR observe a 9% lower risk of T2D (HR: 0.91, 95%CI: 0.85, 0.98) when habitually consuming a cup of coffee compared to those who do not drink coffee (Figure 6-1). There are no comparable studies examining coffee intake with IR GRS and T2D. However, in a study evaluating coronary risk (n=55,685), a long term macrovascular complication of T2D, the authors reported that adopting a healthy lifestyle (a score composed of no smoking, no obesity, regular physical activity and a healthy diet) can reduce the risk of coronary events regardless of genetic predisposition.[236] A healthy diet was defined as adherence to at least half of the following: greater consumption of fruits, nuts, vegetables, whole grains, fish and dairy products and reduced intake of refined grains, processed meats, unprocessed red meats, SSBs, trans fats and sodium.[236] Although in our analysis we did not identify a statistically significant association between other foods or beverages and T2D risk, interactions with GRS were consistently null, thereby indicating a similar interpretation.

Two nominally significant interactions were observed, whereby among those habitually consuming a daily portion of egg and egg products, participants with

higher genetic risk for T2D had higher risk of developing T2D than those with a lower genetic risk ($p_{interaction}$=0.039). This has not been previously reported and the high between-country heterogeneity ($I^2$= 80%), with a possible interaction evident in France, the UK and Spain, warrants further research in these countries in particular. We also found a nominal interaction between SSB intake and those with a genetic predisposition for BMI on T2D risk, on both the multiplicative and additive scale, where the risk of T2D was highest among those with a higher genetic predisposition for BMI and higher SSB intake. This aligns with the previously reported interaction between SSB and BMI GRS on measures of adiposity,[148,220,237] so further research is of interest.

Therefore, currently, the consistent findings of no evidence for interactions between genetic risk for metabolic diseases and foods and beverages with the development of T2D have two main implications. One, these findings provide evidence against possible concerns about genetic fatalism or determinism for T2D, captured by statements like 'our genes are our destiny.'[238] And two, these findings endorse the message that everyone is susceptible to the health consequences of poor dietary habits, such as developing T2D, including those with low overall genetic metabolic risk. This does not preclude future research investigating the interaction between other dietary and genetic exposures.

*Strengths and limitations*

With low statistical power being a barrier to investigating gene-diet interactions in smaller studies, EPIC-InterAct enabled us to examine our research question within a large study with incident T2D that also has dietary and genetic data. This is the first published interaction analyses for T2D which has investigated the interaction between foods and beverages with three GRS, representing key intermediate determinants of and explaining greater variance for T2D. Attempts were made to minimise the possibility of spurious interactions through multiple approaches: i) taking a systematic approach to selecting foods and beverages based on previously reported investigations with T2D, hence avoiding selective testing and reporting, ii) accounting for multiple testing to reduce false positive reporting, iii) confirming that there is little evidence of dependence between the GRS and dietary exposures (Table 6-3)[218] and iv) trying to address potential

confounding including population stratification.[126] The consistent findings of no interaction under both the multiplicative and additive interaction scales further strengthen our null inferences. However, in choosing these 15 foods and beverages, the possibility of interactions with other dietary variables or overall dietary patterns that reflect potential synergistic interplay between different foods and beverages cannot be excluded. A previous analysis in EPIC-InterAct reported no evidence for an interaction between the Mediterranean dietary pattern and T2D GRS,[239] but further investigation in other cohorts and with other dietary patterns are warranted. Although GRS may improve statistical power,[149] some argue that if individual SNPs interact with dietary exposures in opposing directions, this can subsequently reduce statistical efficiency.[150] Lastly, these findings cannot be generalised to populations of non-European descent, but inclusion of eight European countries improved the generalisability to Europeans.

In conclusion, these findings from the EPIC-InterAct Study in eight European countries suggests that the associations of 15 foods and beverages with future risk of T2D are not modified by an individual's genetic predisposition for BMI, insulin resistance or T2D.

# *Chapter 7 Genome-Environment-Wide-Interaction-Study of macronutrient intake on incident Type 2 Diabetes*

This chapter reports the use of a genome-wide approach to test for the interaction between genetic variants and macronutrient intake and risk of developing T2D (relating to objective 5, Chapter 2). To my knowledge, this was the first attempt to investigate this topic using a fully agnostic approach and has raised important issues surrounding method development for such an analysis.

## *7.1 Abstract*

**Background:** to date, no gene-macronutrient interactions in the development of type 2 diabetes (T2D) have been robustly replicated, which invites the opportunity for hypothesis-free methods to potentially identify novel interactions at unexpected genetic loci to improve our understanding of T2D aetiology.

**Methods**: a Genome-Environment-Wide-Interaction-Study (GEWIS) of five dietary macronutrients on incident T2D was conducted, in EPIC-InterAct, a prospective case-cohort study across 8 European countries (N=21,148 with 9,403 cases of diabetes). Using Illumina 660W-Quad BeadChip or Illumina HumanCore Exome chip arrays, approximately 7.8M genetic variants were examined for interaction with the proportion of energy intake from total carbohydrate, total

protein, saturated fat, polyunsaturated fat or cereal fibre intake estimates derived from self-reported food frequency questionnaires. We used Cox regression to estimate the beta coefficients for multiplicative interactions. Genomic inflation factor ($\lambda_{gc}$) was used to evaluate possible systematic bias that may be due to population stratification.

**Results:** no genetic loci demonstrated significant evidence of interaction with the five macronutrients examined in relation to the development of T2D ($p_{interaction} \geq 1.29E-07$, with the threshold for significance<5E-08). Highly inflated test statistics were observed across all findings from the Illumina 660W-Quad BeadChip ($\lambda_{gc} \leq 2.56$). Therefore, results must be interpreted in light of this. Post-hoc analyses were undertaken to investigate the cause of this inflation. The specification of a covariate, BMI as a continuous or categorical variable, greatly contributed to this inflation.

**Conclusion:** further analyses are planned before any conclusions can be made about these null findings. After resolution of methodological issues, it is anticipated that there would be more widespread use of genome-wide approaches for testing gene-diet interactions.

## *7.2 Background*

The literature on gene-macronutrient interactions and T2D, including our null findings, do not highlight any promising examples using hypothesis-driven approaches (i.e. candidate genetic variants or genetic risk scores in Chapters 4 and 5).[209] Therefore, taking a hypothesis-free approach was the next logical step in the systematic interrogation of whether such interactions contribute to the aetiology of T2D, with the potential to identify interactions at unexpected genetic loci.

Thus far, hypothesis-free approaches such as GEWIS have been fairly under-explored, mainly hindered by a lack of statistical power of small sample sizes and the large multiple testing burden. Within the literature examining gene-diet interactions and T2D, only one study has taken a GEWIS approach. This study examined dietary haem iron,[240] whereas none have studied gene-macronutrient interactions.[209] In a relatively modest sized US population (n=4,771), the association between haem iron intake and T2D did not significantly differ by genetic variants, when assessed across 700,000 SNPs as well as specifically with SNPs within iron metabolic pathways.[240] Within cancer epidemiology, GEWIS approaches to studying gene-diet interactions have recently emerged in the literature, with interactions being reported for processed meat,[241] dietary pattern[242] and alcohol intake.[151]

To date, there has been no consensus on the best analytical strategy for undertaking a GEWIS, that is, should it be a 'joint' test with 1 degree of freedom (only examining the interaction term), with 2 degrees of freedom (tests for both gene and gene-environment interaction, proposed by Manning and Kraft),[243,244] or a 'stratified' framework, which compares the main genetic effect between strata of a second exposure.[245] Kraft et al. advocate the 2 degrees of freedom tests for variants with low or no marginal effect.[244] A recent empirical comparison of joint *vs* stratified tests showed a high correlation between the $-\log_{10}$P values ($r>0.75$) for cohort studies, but recommended a joint approach to retain statistical power and if studying low-frequency variants.[245] Some reviews[35,120] have summarised alternative methods, which have often been used in addition to the classic joint

test with 1 degree of freedom, with variable success. This includes 2 step approaches[246] and genome-wide heterogeneity of variance analysis,[247] albeit the latter method cannot determine the exposure causing the heterogeneity.

Therefore, we aimed to characterise gene-macronutrient interactions on T2D risk, at the genome-level, using the most basic joint test with 1 degree of freedom. With this approach we hope to fill a gap in the literature in using GEWIS to examine this research question.

## *7.3 Method*

For the current analyses, the EPIC-InterAct study was used with the main methods already described in Chapter 3. In addition, specific methods for this piece of work are described below.

*Cohort*
Current analyses were based on a subset of EPIC-InterAct with available genome-wide genotyping, dietary data and data on covariates (9,403 cases and 11,745 non-cases, a total of 21,148).

*Dietary data*
As a first GEWIS of macronutrient intake and T2D, five macronutrients (total carbohydrate, total protein, SFA, PUFA and cereal fibre) were prioritised to minimise the multiple testing burden and for feasiblity reasons. Macronutrients were treated as nutrient densities, being either 5% of total energy intake/day or 1g/1000kcal/day for cereal fibre. This was to enable comparability and potential future harmonisation with previously published literature, which often used these units. Total energy intake includes energy from carbohydrate, fat, protein and alcohol intake. The outliers were Winsorised at the 1$^{st}$ and 99$^{th}$ percentile. To reduce possible multicollinearity in the context of interactions, macronutrient variables were centred to the mean.[129]

Rationale for macronutrient selection

1. *Total carbohydrate intake*
2. *Total protein intake*
3. *+ 4. SFA and PUFA intake:*
    - There is no clear association between self-reported intake of total fat, SFA, MUFA or PUFA and T2D.[63,224,248,249]
    - The cardiometabolic literature, however, indicates different directions of associations between the quality indicators of fat intake (e.g. SFA, MUFA, PUFA). SFA and PUFA intake are associated with cardiovascular disease but in opposing directions, therefore investigating total fat intake would mask these effects.[78,81,250] Also, current dietary advice is for replacing SFA with PUFA intake.[208]
5. *Cereal fibre intake:*
    - Meta-analysis show that although both total and cereal fibre are associated with T2D, cereal fibre (RR for T2D per 10g/d: 0.75, 95%CI:0.65,0.86) appears to be the driving factor between total fibre and T2D (RR for T2D: 0.91, 95%CI: 0.87,0.96).[86]

*Genotyping and imputation*

Participants were genotyped on the Illumina 660W-Quad BeadChip (referred to as 'GWAS chip' in this chapter) or Illumina HumanCore Exome Chip (referred to as 'Exome Chip' in this chapter) arrays and imputed to the Haplotype Reference Consortium using IMPUTE v2.3.2. Pre-analysis filtering excluded SNPs with minor allele count <10, imputation accuracy of info<0.4 and Hardy-Weinberg-equilibrium p $<10^{-6}$. An additive genetic model was assumed for each SNP, where the number of risk alleles was 0, 1 or 2 for genotyped SNPs and a continuous dosage ranging from 0 to 2 for imputed SNPs.

*Statistical interaction analysis*

Interactions between each SNP and macronutrient intake on incident T2D were analysed using a Prentice-weighted Cox regression model on the basis of the case-cohort design of EPIC-InterAct, by including the macronutrient, SNP and interaction terms in the model (on a multiplicative interaction scale). Both macronutrient and genetic exposures were entered as continuous variables. Potential confounders relating to both exposures (SNP and/or macronutrient intake) have often not been adequately addressed in past interaction analyses and may induce both type I and II errors, as Keller, VanderWeele, and Sul have empirically demonstrated and discussed.[126–128] Therefore, in balancing this with computational efficiency, we adjusted for age at recruitment (continuous variable) as the underlying time scale, sex (dichotomised), country (categorised: France, Italy, Spain, UK, Netherlands, Germany, Sweden, Denmark), first 3 principal components for population stratification, total energy intake (TEI) (continuous: kcal/day), physical activity (ordinal, treated as continuous: inactive, moderately inactive, moderately active, active), level of education (categorical: none, completed primary school, technical/professional school, secondary school or longer education), smoking status (categorical: never, former or current smoker), sex-specific alcohol intake (categorical: none, light drinking (0.1-6g of alcohol/d), moderate drinking (men: 6.1-24g, women:6-12g/d) and heavy drinking (men>24g, women>12g/d)) based on WHO's definition of 12g of alcohol/drink,[251] and BMI (continuous). Equation 4 below illustrates this, where 'SNP' is the genetic variant and 'Macro' is the macronutrient.

$$f(T2DM) = \beta_0 + \beta_{SNP}SNP + \beta_M Macro + \beta_{SNP \cdot M}SNPxMacro + \sum \beta_C Covariate \quad (4)$$

Chip-specific GEWIS analyses were conducted in *Stata* v14 (StataCorp LP, Texas, USA), using an in-house GWAS programme. The robust command was used to calculate the robust standard error of the interaction term and error covariance, as advised by Voorman et al., to help minimise exposure misspecification under an interaction model.[252] Due to the computational and time intensiveness of these analyses, a specialised high-throughput computing system (called high performance computing) was used. The University of Cambridge is one of six centres in the UK which supports this system.[253] Despite this powerful system,

each GEWIS of macronutrient intake takes approximately three days to run and equates to £2,000 of computing time and energy. Therefore, decisions relating to the analyses were made with consideration of the efficiency of analysis.

Post analysis quality control included exclusion of rare variants with minor allele frequency (MAF) ≤1% and those with missing estimates. Chip-specific results were then meta-analysed using the inverse variance-based method within the METAL software, originally designed to meta-analyse GWAS findings.[254] This resulted in approximately 7.8M SNPs used in our analyses. Between-chip heterogeneity was assessed with Cochran's Q statistic and heterogeneity $I^2$. We used $p_{interaction} \leq 5E\text{-}08$ as the p-value threshold for significance.

Quantile- Quantile (QQ) plots were used to examine if the distribution of $-\log_{10}$ p values for interaction were consistent with the null distribution (except the extreme tail). Deviation from the expected null distribution can suggest large-scale systematic bias due to population stratification or genotyping error. The meta-analysed results (across chips) have been corrected for genomic control lambda.

Chip-specific genomic control inflation lambda ($\lambda_{gc}$) was calculated by the ratio of the observed median chi-squared divided by the expected median chi-squared distribution.[255] A value for $\lambda_{gc}$ of less than one is assumed to be from sampling error alone, whereas a value above one indicates risk of population stratification leading to spurious findings and therefore genomic control correction is usually applied. However, as a rule of thumb in GWAS, a $\lambda_{gc} > 1.1$ is considered unusually high so applying genomic control correction is insufficient. Thus the underlying cause of inflation should be addressed.[256]

If statistically significant interactions were identified, the following sensitivity analyses were planned:

1. Further adjustment for potential dietary confounders
    a. GEWIS for macronutrient intake: dietary total fibre, magnesium, iron, vitamin C, green leafy vegetables, tea, coffee
    b. GEWIS for cereal fibre intake: protein, carbohydrate, saturated fat, monounsaturated fat, polyunsaturated fat intake, magnesium, iron, vitamin C, green leafy vegetables, tea, coffee

2. To assess the possibility of confounding by covariate interaction, interactions between covariates and macronutrient as well as between covariates and SNPs would be modelled simultaneously.[128] For example, this is exemplified in Equation 5:

$$f(T2DM) = \beta_o + \beta_{SNP}SNP + \beta_M Macro + \beta_{SNPxM}SNPxMacro + \sum\beta_C Covariate + \sum\beta_{SNPxC}SNPxCovariate + \sum\beta_{MacroxC}MacroxCovariate \quad (5)$$

   If the strength of the interaction attenuates, distinguishing the covariate responsible will aid understanding of the causal pathway, using the approach proposed by Waldman[257]
3. Verification of gene-macronutrient independence[218]
4. Accounting for isocaloric macronutrient substitution

If interactions were robust to the above analyses, replication would be warranted and annotation of the SNPs of interest would be undertaken.

This methods section reflects the stage of the project that was achieved at the time of writing this chapter. In light of highly inflated $\lambda_{gc}$, post-hoc exploratory analyses were undertaken to determine its cause.

## 7.4 Results

The population characteristics of the EPIC-InterAct subset used for this work were the same as that used for the replication analysis in Chapter 4 (Table 4-3). The main associations between the proportion of energy from macronutrient intake (total carbohydrate, protein, SFA, PUFA and cereal fibre) on incident T2D are shown in Table 7-1. Total protein intake was positively associated with T2D (HR=1.09 per standard deviation; 95%CI: 1.04, 1.14). There were null associations for the other macronutrients examined.

**Table 6-1: Association between macronutrient intake and incident type 2 diabetes (EPIC-InterAct study)**

| Macronutrient intake | N (cases/total) | Mean (SD) | HR (95%CI) per SD |
|---|---|---|---|
| Total carbohydrate (%TEI) | 9403/21148 | 43.85(6.97) | 0.96(0.91,1.01) |
| Total protein (%TEI) | 9403/21148 | 17.03(3.04) | 1.09(1.04,1.14) |
| Saturated fat (%TEI) | 9403/21148 | 13.34(3.34) | 1.01(0.92,1.12) |
| Polyunsaturated fat (%TEI) | 9403/21148 | 5.55(1.81) | 1.02(0.95,1.09) |
| Cereal fibre (g/1000kcal) | 9400/21139 | 8.91(4.92) | 1.00(0.91,1.10) |

Hazard Ratios (HR) for macronutrients (per SD difference) and Type 2 Diabetes (T2D), adjusted for age (=underlying time scale), sex, country, total energy intake, physical activity, education, smoking, sex-specific alcohol categories, BMI
Abbreviations: N- number, SD- standard deviation, HR- hazard ratio, TEI- total energy intake

Among the five macronutrients and more than 7 million SNPs examined, we did not observe any gene-macronutrient interactions contributing to T2D aetiology that were robust to passing the genome-wide-significance threshold (shown by the red horizontal line in the Manhattan plot in Figure 7-1).

The QQ plots for the chip-specific and post-genomic control meta-analysed findings are shown in Figure 7-2. The test statistic inflation ranged from 1.06 to 2.56, with an average of 1.29. There was overall higher test statistic inflation in the GWAS chip analyses, with the PUFA GEWIS most affected ($\lambda_{gc}$: 2.56), followed by carbohydrate GEWIS from the same chip ($\lambda_{gc}$: 1.54). All GEWIS, except that for PUFA and carbohydrate intake were considered at low or moderate risk of systematic type I error.

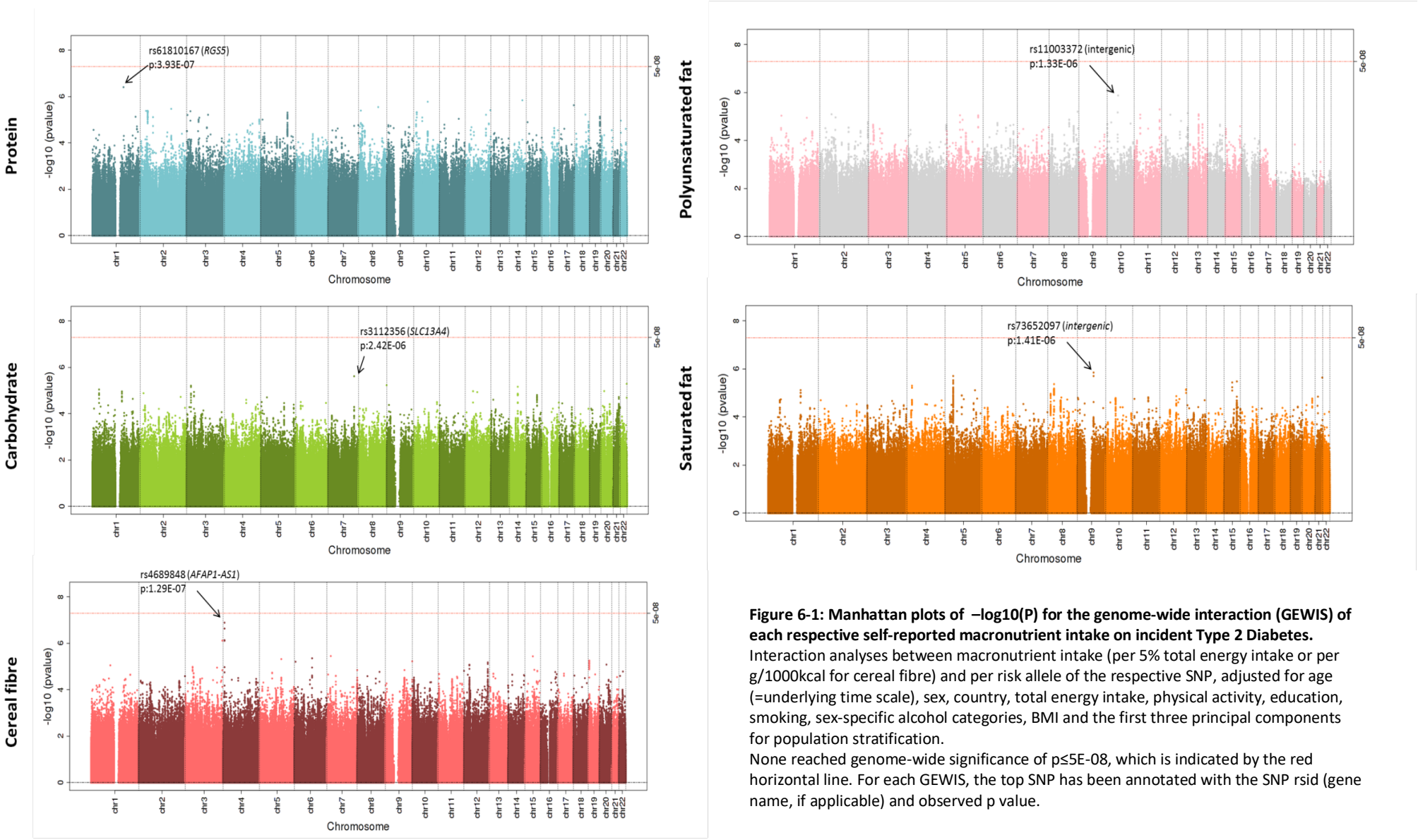



**Figure 6-1: Manhattan plots of –log10(P) for the genome-wide interaction (GEWIS) of each respective self-reported macronutrient intake on incident Type 2 Diabetes.**
Interaction analyses between macronutrient intake (per 5% total energy intake or per g/1000kcal for cereal fibre) and per risk allele of the respective SNP, adjusted for age (=underlying time scale), sex, country, total energy intake, physical activity, education, smoking, sex-specific alcohol categories, BMI and the first three principal components for population stratification.
None reached genome-wide significance of p≤5E-08, which is indicated by the red horizontal line. For each GEWIS, the top SNP has been annotated with the SNP rsid (gene name, if applicable) and observed p value.

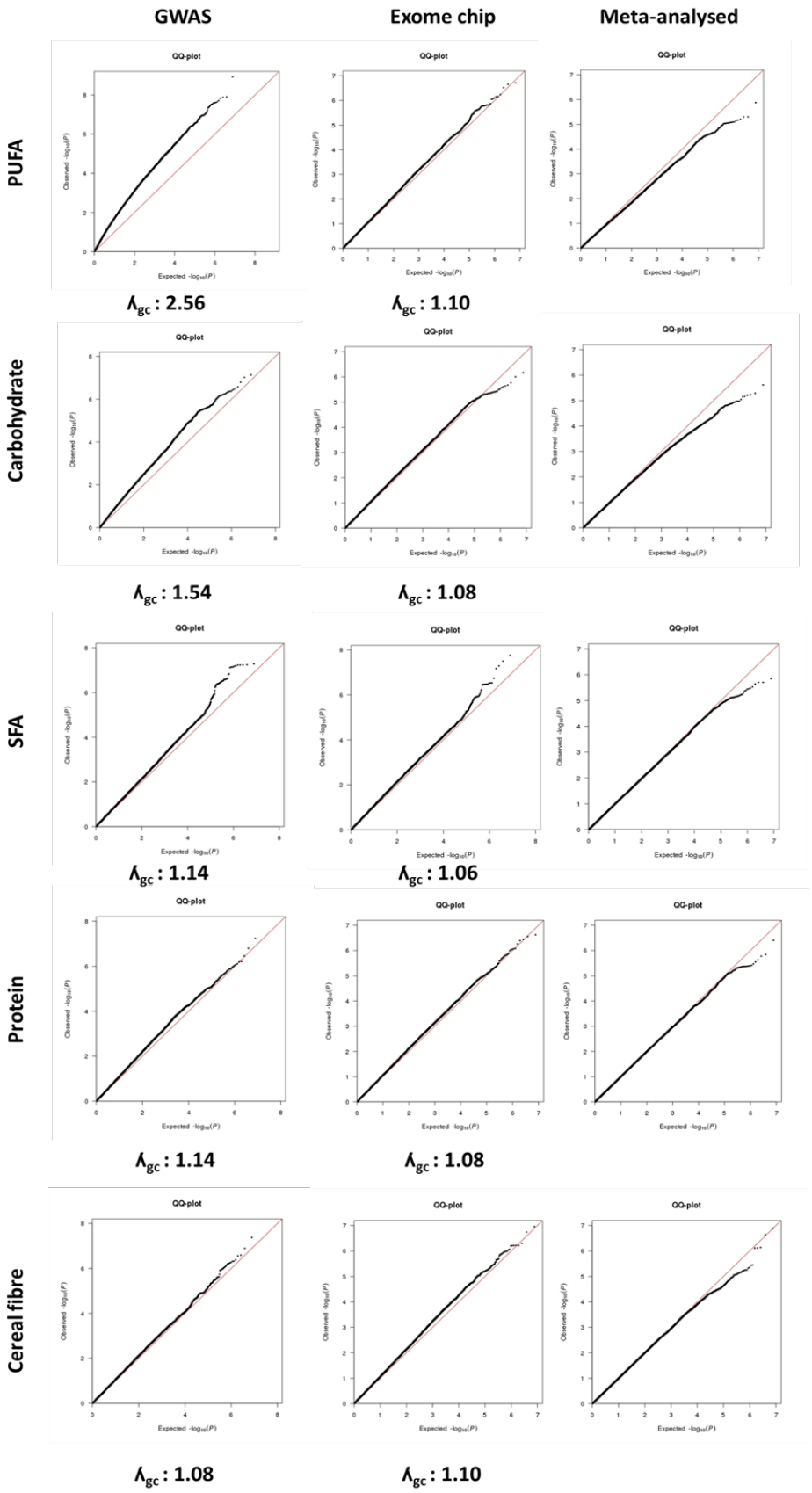



**Figure 6-2: QQ plots of –log10(p) for the genome-wide interaction (GEWIS) of the respective self-reported macronutrient and incident T2D, by genotyping chip**. Meta-analysed $-\log_{10}(p)$ for interaction was undertaken using the METAL software and corrected for genomic control.

*Results from post-hoc exploratory analyses:*

Given highly inflated test statistics ($\lambda_{gc}$) for some GEWIS, a number of exploratory analyses were undertaken to determine the cause of this inflation. The PUFA GEWIS was used as the example because it exhibited the highest inflation ($\lambda_{gc}$: 2.56 on the GWAS chip).

1. No obvious geographical differences in PUFA intake that would have led to the highly inflated test statistics (Figure 7-3 shows the country distributions by genotyping chip).
2. No obvious chip-specific problems (in regards to genotyping or imputation). No inflation was evident in the GWAS chip for either a GWAS (as opposed to a GEWIS) conducted for circulating vitamin C (Zheng et al., unpublished) or a GEWIS conducted for BMI and incident T2D, in EPIC-InterAct (Sharp et al., unpublished) (Figure 7-4). This may suggest that the inflation observed in our analysis is specific to GEWIS of dietary factors and incident T2D, if not specific to macronutrient intake. Findings from the above two checks suggest that there were no obvious differences in participant characteristics between chip arrays that would have importantly influenced the findings.
3. Potential multicollinearity was examined for the PUFA GEWIS on the GWAS chip. Firstly the effect of reducing the number of covariates of the multivariable model from 12 to 8 covariates (excluding physical activity, smoking, education and alcohol intake) was examined. This increased the $\lambda_{gc}$ from 2.56 to 3.56 (Figure 7-5), with the explanation of this observation unclear. In suspecting that the prospective nature of the Cox regression may be a reason, all covariates except for age as the underlying time variable was dropped from the model to assess the influence of age, however, the inflation disappeared ($\lambda_{gc}$: 1.02) (Figure 7-5). This confirmed that age was not the problem. To assess how much each of the 8 covariates contributed to this inflation the analysis was performed 5 times, each time excluding one covariate (all 3 principal components were dropped simultaneously) from the model. The model without BMI abolished the inflation ($\lambda_{gc}$: 0.985), which indicates that BMI had the largest influence (Figure 7-6). For computational efficiency, analyses were performed using

only chromosome 22, after confirming that the inflation existed uniformly across all chromosomes. Given that there was no notable difference in the distribution of BMI between genotyping chips, we concluded that at least for the PUFA GEWIS findings, conducted on the GWAS chip, the inclusion of BMI in the model contributed to high inflation.

4. Whether the covariate, BMI, may have been misspecified. It was observed that treating BMI as a categorical rather than as a continuous variable reduced the inflated test statistic. Figure 7-7 shows the GEWIS for PUFA and carbohydrate intake, respectively. This was based on analyses adopting the most adjusted model, where BMI was treated as WHO cut offs (<25, 25-29, ≥30kg/m$^2$), only on the GWAS chip and only for chromosome 22.

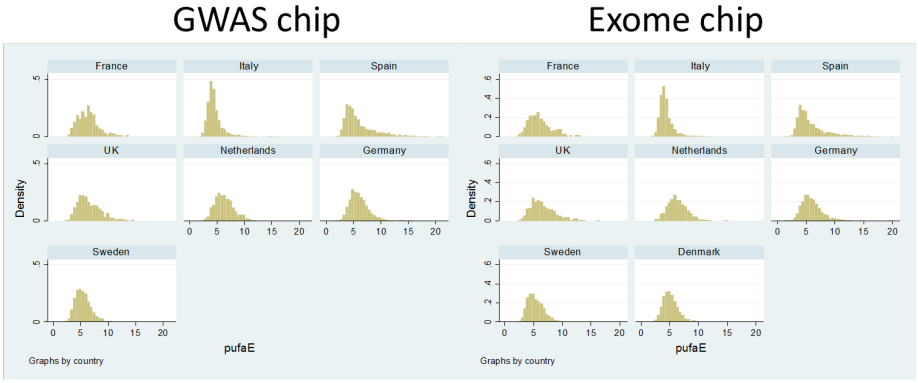



**Figure 6-3: PUFA intake distributions within each country, by the GWAS and Exome Chip**

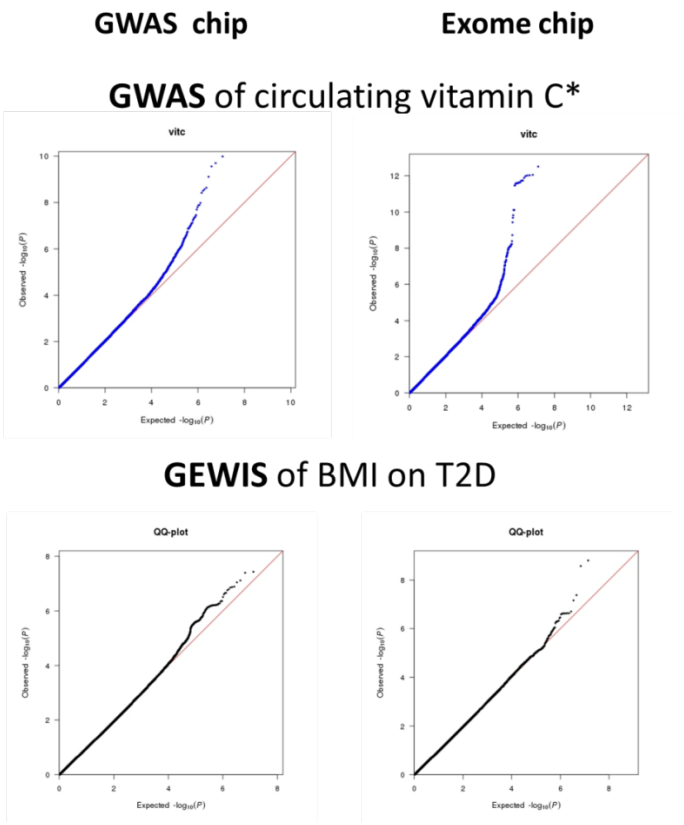



**Figure 6-4: Genotyping chip specific QQ plots of $-\log_{10}(p)$ for a) the GWAS of circulating vitamin C (Zheng, unpublished) and b) GEWIS of BMI on incident T2D (GWAS chip N=9,100, Exome chip N=13,078) (Sharp, unpublished).** * indicates in subcohort only (GWAS chip N=3,521, Exome chip N=6,504).

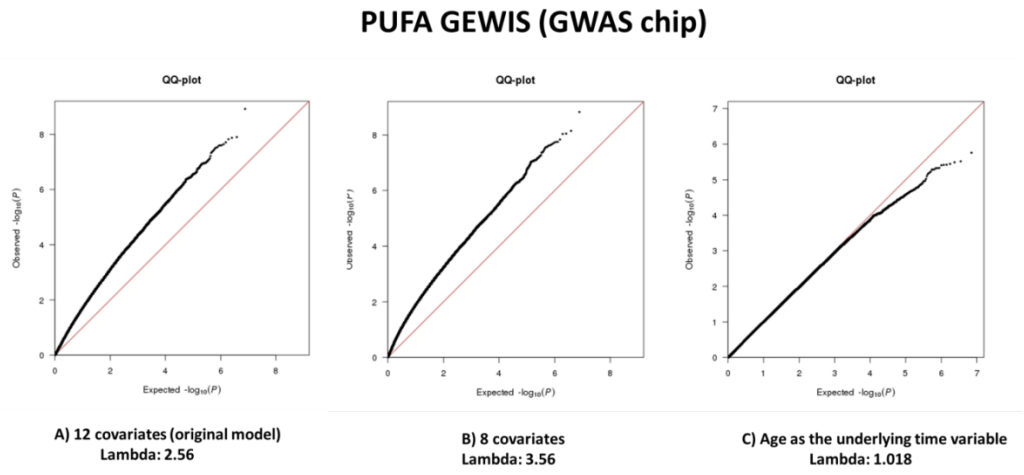



**Figure 6-5: QQ plot of $-\log_{10}(p)$ for interaction from the PUFA GEWIS on the GWAS chip only.** Shows that reducing the number of covariates from 12 to 8 increased the inflation (A compared to B) and that age was not inducing the inflated test statistics (C). PUFA GEWIS adjusted for A) age (underlying time scale), sex, country, first 3 PC for population stratification, total energy intake, physical activity, education, smoking, sex-specific alcohol intake, BMI; B) age (underlying time scale), sex, country, first 3 PC for population stratification, total energy intake, BMI; C) age (underlying time scale).

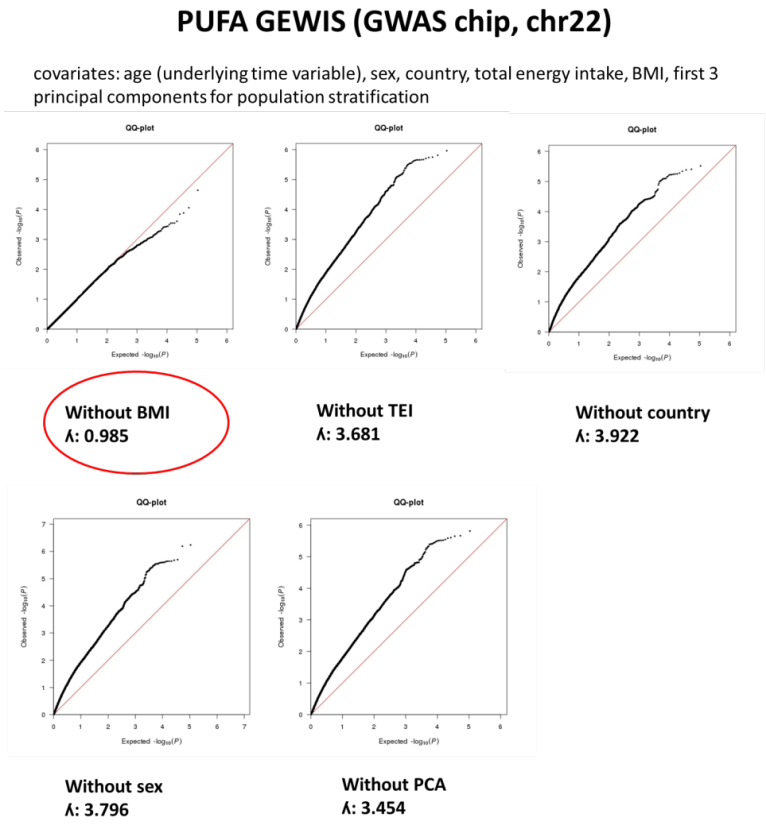



**Figure 6-6: QQ plots of $-\log_{10}(p)$ for interaction from the PUFA GEWIS**, only on the GWAS chip and on chromosome 22. Each of the above covariates was dropped one at a time. Abbreviation TEI: total energy intake, PCA: principal components for population stratification

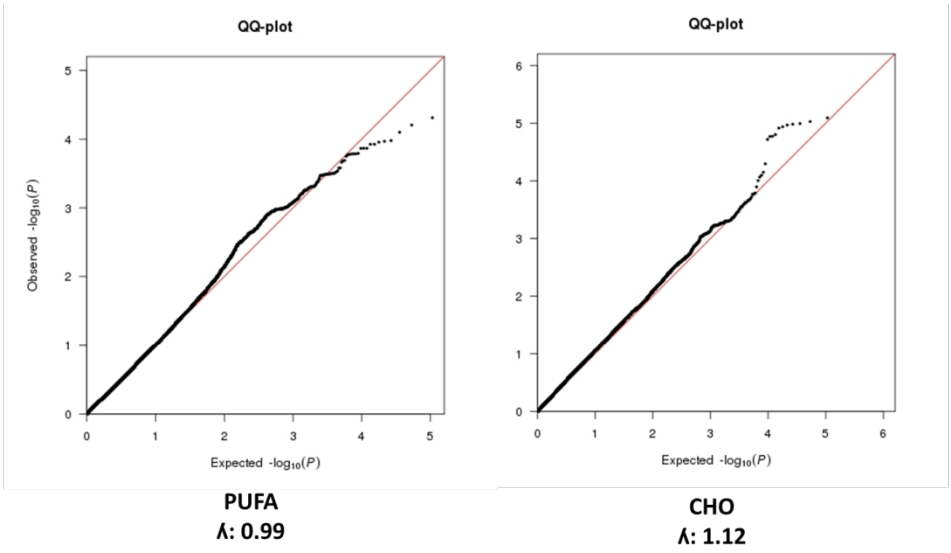



**Figure 6-7: QQ plots of $-\log_{10}(p)$ for interaction from the PUFA and carbohydrate GEWIS**, only on the GWAS chip and on chromosome 22, under the most adjusted model using categorised BMI as a covariate. Analysis adjusted for age (underlying time variable), sex, country, TEI, first 3 principal components for population stratification, BMI, physical activity, education, smoking and sex-specific alcohol intake.

## 7.5 Discussion

To our knowledge, this was the first macronutrient intake GEWIS for incident T2D, using over 7 million SNPs, examined in over 21,000 participants. No interactions at genome-wide-significance were identified. For some analyses, the high genomic inflation was in a large part due to misspecification of a covariate (BMI). This methodological issue has therefore shaped future plans for this piece of work and has raised an important consideration for future GEWIS of dietary intake.

The high degree of inflated test statistics observed for the PUFA and carbohydrate GEWIS findings resulted in an evaluation of the suitability of QQ plots to assess systematic type I error in the setting of a GEWIS. It also led to identifying a gap in the literature for methods in evaluating reasons for this beyond population stratification. Although assessing QQ plots and the degree of inflation of the median test statistic is well accepted in GWAS of a single exposure, Voorman and colleagues demonstrated that, in the GEWIS setting high inflation within a QQ plot could reflect systematic underestimation of variability in interaction effect estimates across the genome, therefore resulting in spurious QQ-plot inflation even when there is no problem with population substructure.[252] Theoretically, gene-environment interactions are more prone to violating the model assumption that the mean value of the outcome is truly linear in the coefficients of the exposure variables. This is because, in nutritional epidemiology, we have seen that many dietary exposures show non-linear associations with disease, including the U-shaped association between alcohol intake and T2D.[251] Therefore, Voorman et al., proposed modelling robust estimates of variance to prevent misspecification. In all the current analyses, this was done, yet high inflation was still evident. In a GEWIS of BMI and T2D within the Nurses' Health Study, Cornelis and colleagues simulated four methods for specifying the main exposure (BMI) which resulted in different levels of inflated test statistics.[205] They concluded that treating their main exposure (BMI) as a linear continuous variable led to variances of the interaction β to be underestimated and therefore inflated test statistics. This was rectified by modelling BMI as polynomial, dichotomising it or using a robust variance estimator, as suggested by Voorman et al. In our

case, categorising our main exposure (proportion of energy from PUFA intake) and using a robust variance estimator did not fully eliminate the inflated test statistic ($\lambda_{gc}$: 1.32), in the context of a larger number of T2D cases (N=9,403) than that available in the Nurses' Health Study (N=2,199 cases of T2D). This demonstrated how currently proposed methods for controlling test statistic inflation in the GEWIS setting,[205,252] if assuming that misspecification of exposures may be a contributing factor, is insufficient. In the current analysis when the covariate, BMI, was subjected to the same methods adopted by Cornelis et al., a decrease in inflation for both the PUFA and carbohydrate GEWIS was observed (Figure 7.7). Therefore, although further methodological developments are warranted, the findings indicate that future GEWIS of dietary and/or macronutrient intake should apply robust variance estimator during analyses and if QQ plots indicate inflated test statistics, researchers should evaluate how the main exposure(s) and covariate(s) are specified before concluding that population stratification may have biased their findings.

The differences observed for findings between the two genotyping chips is difficult to explain and the possibility of collider bias contributing to the role of BMI is challenging to assess in this setting.[258]

The question about whether it is too early to establish a fixed threshold for rejecting the null from GEWIS analyses, is still under debate. The conflict between a) raising the type II error rate by using stringent thresholds to determine whether an interaction is due to chance and b) raising the type I error rate when using less conservative approaches, is difficult to balance. In a meta-analysis of 22 candidate gene-based interaction studies, the authors demonstrated that *NAT2* slow acetylates are more susceptible to the adverse effects of cigarette smoking on bladder cancer risk because they have a reduced detoxification capacity for aromatic monoamines, found primarily in tobacco smoke.[103] This was identified at a p value for interaction much higher than the genome-wide-significance threshold used in the GEWIS setting ($p_{interaction}$=0.008).[103] Hence, this demonstrates a risk for type II error when undertaking a GEWIS because biologically plausible interactions, including this, may be overlooked when a stringent threshold for rejecting the null hypothesis is applied. However, given the likelihood of 'winner's curse,' and the potential for

spurious interactions in the context of many limitations associated with observational epidemiology that could lead to misinterpretation at the public health interface, a more conservative approach to investigating interactions by using the genome-wide-association threshold ($p_{interaction}$ ≤5E-08) was chosen in the current analyses. This appears to be current practice for GEWIS investigating various outcomes.[151,241,259] Moreover, enlarging the current dataset by replication in comparable studies and then pooling the findings, particularly of the interactions that were at sub-genome-wide-significance threshold ($p_{interaction}$ ≤1E-06), may offer further insights.

*Strengths and limitations*

This current work has tried to address some of the known methodological concerns relating to GEWIS approaches, including potential confounding for both exposures and had planned a thorough set of sensitivity analyses to minimise the likelihood of type I error. Additionally, an important methodological issue not previously reported in the GEWIS literature has been highlighted and may pertain specifically to GEWIS of dietary exposures, if not to macronutrient exposures. However, it is questionable whether there was sufficient power to detect interactions with the small magnitude of effects that were observed. Appendix I lists five SNPs for each macronutrient GEWIS that achieved the lowest p values for interaction. Among these SNPs with a MAF≥10% the interaction beta ranged from 0.07 to 0.62, with all but one SNP demonstrating an interaction beta of ≤0.22. Our analyses had <80% power to detect an interaction beta of ≤0.25, assuming an MAF≥10% within EPIC-InterAct (N=9,403 incident T2D cases, see Figure 7-8).[156] Despite this, EPIC-InterAct is a large prospective study with incident T2D (with over 5 times the number of T2D cases as other large prospective observational cohorts)[142] and both genetic data and measures of self-reported macronutrient intake. It would be worthwhile to consider enlarging the discovery dataset by meta-analysing with comparable studies to improve statistical power. This would require considerations for harmonisation of the genetic and dietary exposures as well as the analysis approach.

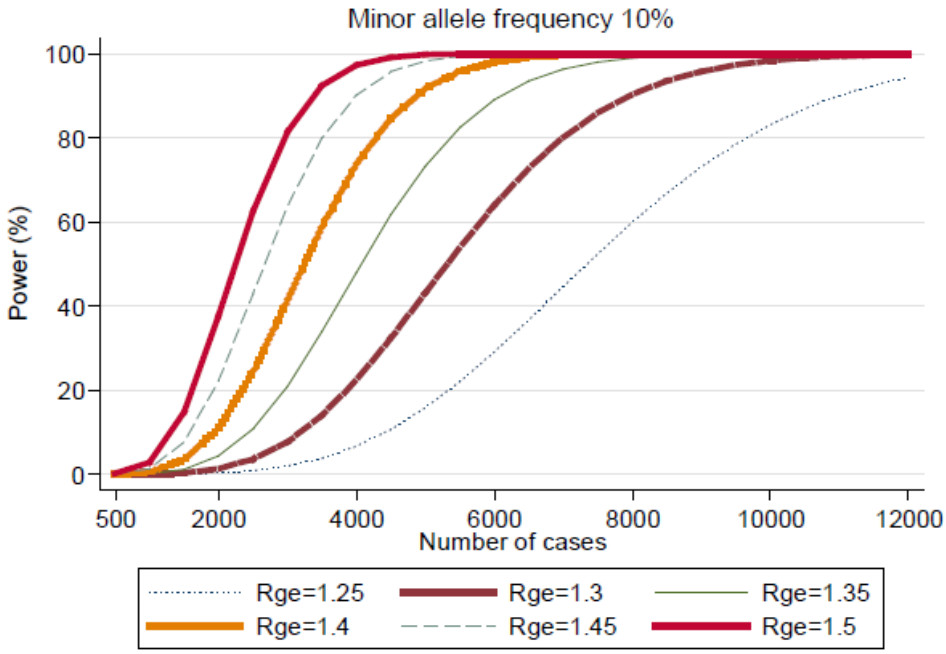



**Figure 6-8: Power calculations for detecting interactions between genetic variants and environmental exposures within EPIC-InterAct** (software: QUANTO based on assumed population risk of 5%, MAF of 10% and normally distributed environmental exposure).[156] Rge: relative risk (magnitude of effect) for the interaction between genetic variants and environmental exposure.

*Future plans*

Given that a practical method to address the inflation observed for the PUFA and carbohydrate GEWIS would be to treat the BMI covariate as a categorical variable, this is the plan for when these two macronutrients will be re-analysed.

As previously mentioned, interrogating interactions at sub-genome-wide-significance ($p_{interaction}$ ≤1E-06) via replication in comparable studies and then meta-analysis, would be a next step. From the meta-analysis, an option to better understand the biology of statistically significant interactions ($p_{interaction}$ ≤5E-08) could be to filter by genes of biological relevance to the macronutrient of interest and/or T2D pathophysiology.[260] To begin with, a focused list of genes within a pathway for each respective macronutrient will need to be identified. This can be informed by genetic variants known to cause protein change (nonsynonymous) or have been shown to regulate gene expression using databases such as the Genotype-Tissue Expression (GTEx).[261] However, some considerations in using

such an approach include the complexity in gene regulation (i.e. whether they are *trans* or *cis-regulated* genes and if there may be interaction between them) and tissue specificity of gene expression. The selection of appropriate pathways can also be related to genes involved in inherited metabolic disorders of nutrient metabolism, for instance the mutation in *GALT* which encodes for the enzyme involved in galactose (a carbohydrate) metabolism which results in the condition called galactosemia.[262] Another resource is the Kyoto Encyclopedia of Genes and Genomes (KEGG), a database of biochemical pathways highlighting enzymes and other cofactors that are involved in the digestion, absorption, metabolism, distribution and elimination of macronutrients.[263] The genetic variants identified from this approach are key candidates to prioritise for future gene-macronutrient interaction testing under various conditions (e.g. different population groups) and methods (e.g. using a recall by genotype RCT study design).

This first attempt to examine gene-macronutrient interactions across the genome has highlighted several methodological issues that warrant further investigation and development to guide future GEWIS research applied to dietary intake. Although much has been learnt from GWAS studies, it must be borne in mind that investigating interactions (as distinct from marginal genetic effects in the presence of interactions) involves the convergence of two fields of epidemiology (nutrition and genetics). Therefore, decisions about multiple testing corrections, selection of covariates, interpretation of QQ plots, and how exposure and covariates are specified, are some that require not only considerations of the conventions taken in both fields but perhaps also the development of novel methods and practices specifically targeted to GEWIS in the context of particular exposures, such as diet. Although this piece of work remains on-going, several gaps in understanding have been identified and warrant further methodological research to help with advancements within this field.

# *Chapter 8 General discussion, implications and conclusion*

Within my PhD, the overall aim was to understand the contribution of gene-diet interactions to the aetiology of type 2 diabetes (T2D). In particular, I focused on potential interactions with macronutrient intake (Chapters 4, 5 and 7). I have also included investigation of the potential interactions with foods and beverages (Chapter 6). In the chapter-specific discussions, the potential of this leading to genotype-target nutritional interventions to prevent T2D has been evaluated. I took a systematic approach in comprehensively investigating gene-macronutrient interactions and T2D using three methods (candidate genes, genetic risk scores and a genome-environment-wide-interaction study). The following discussion will bring everything together.

## *8.1 Summary of key findings*

In chapter 4, a systematic review of the literature on gene-macronutrient interactions and T2D was conducted, where 13 publications that reported 8 unique statistically significant interactions were identified. These were between one of the following macronutrients: carbohydrate, fat, saturated fat, dietary fibre, and glycaemic load derived from self-report of dietary intake and circulating n-3 polyunsaturated fatty acids with genetic variants in or near the following genetic loci: *TCF7L2*, *GIPR*, *CAV2* and *PEPD* ($p_{interaction}<0.05$). Some of the key limitations of the published literature on this topic included a lack of

reported replication, insufficient correction for multiple testing, inadequate adjustment for potential confounders relating to both genetic and dietary exposures and that no studies accounted for isocaloric macronutrient substitution.

In chapter 4, it was also demonstrated that none of the interactions reported in the literature could be replicated in a large prospective cohort study of Europeans with incident T2D cases (N=21,148 with 9,403 T2D cases) with a comparable population to those of the published studies. This discrepancy in findings likely reflects either true heterogeneity between study populations, methods and/or analyses, or the possibility of spurious interactions identified in previous publications because no studies confirmed their findings via replication. It is probable that chance, confounding and/or biases within the published literature may have contributed to possible false positive findings.

From the systematic review, I identified two key research gaps: i) the need for replication of interactions and ii) there is a paucity of studies investigating gene-macronutrient interactions using non-candidate gene methods, such as using a GRS and GEWIS.

Chapter 5 extended these findings when gene-macronutrient interactions for incident T2D were examined using three GRS (based on meta-analyses of GWAS for BMI, IR and T2D) and no interactions were identified. This was consistent with the null findings reported in Chapter 4.

In chapter 6, using a similar method as for Chapter 5, the investigation was extended to include potential interactions between 15 foods and beverages and these three GRS (BMI, IR and T2D). Again, no significant evidence for interactions were observed for foods or beverages.

Finally, in continuing to examine gene-macronutrient interactions across the spectrum of methods, at the hypothesis-free end, a GEWIS was performed. No statistically significant interactions were identified at the genome-wide significance level for total carbohydrate, protein, saturated fat, polyunsaturated

fat and cereal fibre intake on incident T2D. This work highlighted a number of challenges using GEWIS to examine interactions with macronutrient intake, including the specification of main exposures as well as covariates and which p value threshold to use to reject the null hypothesis.

Taken altogether, this comprehensive approach to the investigation of gene-macronutrient interactions and incident T2D demonstrates that such interactions do not appear to significantly contribute to the aetiology of T2D. Therefore, the current results support universal rather than genetically targeted approaches for dietary advice on macronutrient consumption for the prevention of T2D.

## *8.2 Overall strengths and limitations*

Specific strengths and limitations pertaining to each project within this PhD have been discussed within their respective chapters. The focus of this section is to highlight the overall strengths and limitations.

### Strengths

To my knowledge, the work conducted within this PhD represents the most comprehensive and systematic investigation of the contribution of gene-macronutrient interactions in the aetiology of T2D. We have taken advantage of high-efficiency computation in large genetic datasets to examine this from both hypothesis-driven and hypothesis-free angles. The prioritisation of candidate genetic variants selected for hypothesis-driven approaches was underpinned by a systematic review of the literature and biological plausibility.

A further strength of this work was the opportunity to examine these research questions within EPIC-InterAct, which currently includes a large population of incident T2D cases. This played an important role in maximising statistical power; often a methodological barrier in studying gene-environment interactions. Having had eight countries represented across Europe allowed the investigation of the consistency between countries. It also supported a consistent method for exposure and outcome assessment and analysis across countries, which

minimised heterogeneity. Therefore, these findings can be generalised across Europe.

An attempt was made to reduce the limitations evident in previous studies that investigated gene-macronutrient interactions. This included examining both multiplicative and additive interactions where appropriate (i.e. interaction between GRS and foods and beverages), addressing confounding of both exposures when examining 'interaction' (as opposed to effect modification), addressing multiple testing concerns, transparency in reporting all findings, and where appropriate, reporting both effect estimates by strata of the third variable as well as the p-value for interaction.

Finally, the last project using a GEWIS approach presents an example of challenges that arise from method development of this technique applied to nutrition. Hopefully, the learnings from this work can guide future research using this method.

**Limitations**

The current findings and conclusions are limited by the scope of our exposure selection and should be interpreted with this caveat in mind. It is impossible to have examined every genetic locus and dietary exposure to make a broad generalised conclusion that gene-diet interactions do not contribute to T2D aetiology. Given that intermediate traits for T2D such as BMI and glycaemic markers are more responsive to lifestyle change in part due to their shorter time of effect, it can be hypothesised that gene-diet interactions could have more noticeable effects on these traits. Although examining these intermediate traits would be useful, it was outside the scope of this PhD and thus deserves future attention. Additionally, the majority of the analyses within this PhD were conducted within one study: EPIC-InterAct. Although a phenotypically rich and large study, it does raise issues relating to generalisability and will be discussed in section 8.3.5.

## *8.3 Interpretation of gene-macronutrient interaction findings*

Within the field of cardiometabolic gene-environment interactions, some people have posited that null findings or a lack of replicated interactions could reflect methodological limitations of observational epidemiology. This refers to possibly bias and/or confounding reducing interaction effects so that they are either underestimated or undetectable, even within large population studies.[118] The findings from this PhD indicate that there are no significant gene-macronutrient interactions contributing to the aetiology of T2D. However, next, an evaluation of the internal validity of these findings will be provided to assess whether alternative explanations may explain these in light of epidemiological issues such as chance, bias and confounding. This will then be followed by an assessment of the finding's external validity or generalisability.

### *8.3.1 Chance and statistical power*

The main study used within this PhD was EPIC-InterAct, which was a subset of people from the general population at the time period between recruitment and follow-up (see Methods Chapter 3 for details). However, sampling variability can influence how accurate and precise the findings are for the general population, within this subset of the population. In order to make inferences about the general population from this study, the degree to which chance affected the findings must be determined.

The degree of chance findings can be quantified using p-values and confidence intervals from hypothesis testing. The p-value is defined as the probability of the observed result occurring purely by chance, assuming that there is truly no association between the exposure and outcome. Within this thesis, various p-value thresholds have been used to determine the significance of findings. For replication conducted in Chapter 4, the conventional p-value threshold of 0.05 was used for independent replication. This is an arbitrary cut off meaning that the result of 1 in every 20 tests would be wrongly classified as a real interaction when in fact it is due to chance (type I error or false positive findings), hence no matter how small the cut off is set there is always the possibility of chance. A criticism of published research on gene-environment interactions has been the large number

of tests conducted and hence the high rates of spurious findings, particularly with selective reporting of only those that pass p-value thresholds of 0.05. Therefore, in trying to overcome this, an attempt was made to lower the probability of false positives. Where multiple independent hypothesis tests were conducted, the statistical p-value (e.g. nominal p-value) for interaction was reported but the threshold for rejecting the null hypothesis of no interactions was corrected for multiple testing using Bonferroni correction or the False Discovery Rate, where indicated. However, because some exposures were correlated (e.g. Pearson correlation was 0.7 between total and animal protein intake reported in Chapter 5), this violates the assumption of independence under Bonferroni correction and therefore an alternative 'effective number ($M_{eff}$) of independent tests,' was calculated to minimise the risk of overlooking possible true hits (type II error or false negative findings).[203] This method uses principal component analysis for the exposures to calculate the $M_{eff}$ of independent tests. This number was then used to correct the p-value threshold for significance. For example, in Chapter 5, we investigated the interaction between 3 GRS and 12 macronutrients and found that no interactions were considered statistically significant ($p_{interaction} \geq 0.20$). This was based on the p-value threshold of <0.0015 (0.05/33 is the estimated $M_{eff}$ number of independent tests). This would mean that less than 2 in every 1000 tests would be a false positive finding. For our GEWIS, we used the conventional genome-wide-significance p-value of 5E-08 because it was hypothesis-free and the multiple testing burden was assumed to be similar to a GWAS. Although our findings were all null, based on our large sample size and significance threshold that accounted for multiple testing, any detected interactions would not have been influenced by chance.

How precise the null estimates are is influenced by sample variability, which inherently decreases when increasing the sample size of the study from which the inference is based. Therefore, having used a study with the largest sample size to date to examine gene-macronutrient interactions on T2D incidence, this improves the confidence in making reliable inferences. The 95% confidence interval around our effect measures have also been provided as this indicates the range of values within which the true effect measure lies, with 95% assurance of

accuracy. The narrower the confidence intervals, the greater that estimates are precisely reflecting the true estimate.

Concerns have been raised about the overly-conservative correction that the Bonferroni method enforces, with some suggesting the use of an alternative method being the 'False Discovery Rate.' This estimates the proportion of errors due to falsely rejecting the null hypothesis of no interactions (proportion of false positives you are willing to accept).[264] However, given the currently high level of scepticism within the field of gene-environment interaction regarding spurious findings, a false positive interaction finding may inflict more negative implications for public health and research than a false negative. This was the rationale for choosing a more conservative correction method for the majority of the analyses.

Within genetic epidemiology, where effect sizes of individual SNPs are modest, studies with large sample sizes are usually needed to detect whether the effect truly exists (preventing a false negative). This problem is further inflated when examining the interaction between genetic variants and another exposure, often requiring >10,000 individuals if the expected magnitude of the effect is modest (interaction HR<1.2) and alleles are commonly distributed in the population (MAF>10%).[121] This has been a concern for the interaction field in general.[35,119] The statistical power of a study is the probability of rejecting the null hypothesis if one truly existed (1-β) and where β is the probability of making a false negative error. In a study without adequate power to detect an interaction, a null finding may not necessarily mean a true lack of interaction effect, but rather it may indicate that the signal was not strong enough to be detected within the sample size available. In the first analysis, where an attempt was made to replicate the interactions previously reported (Chapter 4), EPIC-InterAct was the study with the largest number of incident T2D cases (>5 times that of published studies) to have investigated this research question. Therefore, the analyses conducted had >80% power to detect 6 of the 8 reported interactions. For the interaction analyses using GRS, which included a combination of SNPs, no power calculations were performed because i) there have not been any previously published studies to determine the expected magnitude of effect and ii) currently there is no guidance on how to specify the minor allele frequency (MAF) of a

combined GRS. This is a limitation that is acknowledged. In the GEWIS, a detailed discussion has already been made about the likelihood of the study being underpowered (please see strengths and limitations section of Chapter 7).

### *8.3.2 Bias*

Bias refers to a systematic error within a study which results in an incorrect estimate of the association between exposure and outcomes. Next, different forms of bias that have been considered throughout this PhD will be discussed.

#### *Measurement bias; regression dilution bias and misclassification*

To minimise time, expense and participant burden within large epidemiological studies, a food frequency questionnaire (FFQ) is often used to assess self-reported dietary intake, as was the case with EPIC-InterAct. For example, in the UK centres FFQs are limited to 130 items, meaning they are unable to capture all intakes and therefore leading to imprecision in estimating macronutrient intake. Such imprecision can bias the interaction effect towards the null (regression dilution bias) and/or misclassify individuals into categories of exposure. For example, within the replication work described in Chapter 4, it is unclear how much such a bias may be an alternative explanation for the discrepancy that was observed between the null findings from the current work and the previously reported interaction for dietary fibre and *TCF7L2*. In Chapter 4, it was demonstrated that the differential measurement bias for dietary fibre intake across countries did not materially influence the interaction findings, however, the extent to which this impacts on other interactions is unclear. Also, although validation studies have shown that FFQs estimate macronutrient intake well, these studies have not examined quality markers of macronutrient intake including dietary fibre. All self-reported dietary intake, which can also include repeated 24 hour recalls and food diaries, may be subject to recall and social desirability bias. Moreover, in the estimation of macronutrient intake, conversion of food into nutrient data by the use of food composition databases can introduce another layer of error. Especially since interaction effects are particularly sensitive to loss of statistical power due to measurement bias,[122] future gene-diet

interactions should use objective biomarkers of intake, when available and feasible. However, currently a limited set of nutritional biomarkers is available for macronutrient intake and much work is needed to develop, validate and apply biomarkers of diet in nutritional epidemiology. Although statistical methods may be used to correct for measurement bias, as discussed in Chapter 4, using a more precise tool to begin with is ideal.

### *Bias in modelling*

A statistical consideration, already mentioned in the methods chapter, is that how dietary exposures are modelled (either continuous or categorical), can influence the interaction effect and its variance.[205] To minimise loss of power through categorisation, dietary exposures were specified as continuous variables for all analyses except the attempt to replicate previously published interactions (chapter 4). The consistency of approach for exposure specification also ensured that significant interactions were not identified as a result of data dredging.

### *Dissemination bias*

Dissemination bias includes publication bias, selective reporting and language bias and is known to be prevalent in the interaction literature.[265]

It is suspected that non-significant interactions are often unpublished, in the context of many post-hoc interaction analyses and that many reported interactions have not been replicated.[142] This is therefore a limitation of the literature underpinning the systematic review, described in Chapter 4. This practice is not specific to the field of interactions but has implications for interpreting the overall evidence for gene-diet interactions if the totality of evidence is skewed towards positive findings and lead to a wasteful use of resources in conducting research that has already been conducted but not reported. Moreover, from conducting the systematic review, a high level of selective reporting was evident, particularly if interactions were null. Throughout this PhD, I have strongly advocated for publishing all findings, regardless of their results and have included the summary estimates, not just p values for interaction, in supplementary tables as a possible resource for future studies. However, editor bias in accepting positive findings is a well-recognised problem contributing to publication bias that needs to be separately addressed.[266]

Tabery offers a fascinating perspective on how our unconscious biases, which underpin our beliefs and values placed on finding interactions, could influence our expectations of what we find and this, in effect, may result in dissemination bias. In his book, 'Beyond Versus: The struggle to understand the interaction of nature and nurture,' he examined the debate regarding the existence of interactions from a philosophical perspective.[267] This debate has played out since the 20th century and began with RA Fisher who believed in the independent contribution of genetics and environment, whereas Lancelot Hogben advocated for the interdependence of gene and environmental factors on the variation of phenotypes.[268] More contemporary yet equally controversial examples of a similar debate includes the potential existence of the interaction reported by Caspi and Moffitt about whether life stress and depression risk may be moderated by a genetic variation near the serotonin transporter (5-HT T) gene.[104] The commonality between these debates lies in the two perspectives given, which include those who take the a) variation-partitioning approach and believe interactions are purely statistical so need to be eliminated (sceptics) *vs* b) the view that interactions enable understanding of mechanisms and developmental processes that should be understood (advocates).[269] These two distinct perspectives to how scientists study interactions and their extremely different reasons in doing so may, therefore, contribute to how they interpret and share their findings with the scientific community. For variation partitioners, the research question is about how much variation can be explained in a population by the specific exposures and therefore interactions may present an obstacle. Whereas for mechanism-elucidators, interactions aid in understanding the developmental process.[269] What I take from Tabery's explanation is that subconscious biases can influence analysis, interpretation, synthesis of data and perception of what is seen as 'true' in this field despite that quantitative findings are generally perceived as objective. This can also be understood under the framework of social constructivism, where each individual perceives the same experience through a different lens which is influenced by their past experiences.[269] This can lead to dissemination bias and influence the entire process of conducting research. So, it may, therefore, be valuable for researchers to incorporate reflective practices when studying interactions, such as using a

reflective journal to acknowledge their experiences and biases to promote transparency.[270] In setting an example, I have declared my biases below:

> *As a dietitian, I came into this research with the aim of trying to objectively evaluate whether services that claim to provide 'personalised nutrition,' based on an individual's genetic profile are scientifically valid. This led me to study gene-diet interactions. I cannot deny that I was once excited by the prospect of providing more effective dietary interventions through knowledge of someone's genetics. However, my main motivations have always been to promote health and protect the interests of the public. In addition, given that commercial companies have been strong advocates for personalised nutrition, I am aware that I leaned more towards being a sceptic rather than an advocate for gene-diet interactions. To counterbalance this, throughout my PhD I have made both mental and written notes to help me stay mindful of this bias, so I hope that my 'trained' objectivity is apparent within this thesis. I believe everyone is a victim to some form of subconscious bias but it is how mindful we are of this that determines how optimally we treat the information within our power.*

#### *Gene-environment dependence*

It has been recognised that under gene-diet dependence, a statistical interaction can be found for a genetic marker even if there is no interaction between the causal variant and the dietary exposure. Dudbridge and Fletcher explain that this may occur under three scenarios, including by mediation, pleiotropy and confounding.[218] In the analyses conducted within this PhD (Chapter 5 and 6), little correlation was observed between the GRS and dietary exposures. This therefore minimised potential gene-diet dependence leading to possible spurious interaction findings. Understandably, this was not feasible for the GEWIS analyses (Chapter 7) given the millions of SNPs examined but could be undertaken as a sensitivity analysis.

### *8.3.3 Confounding*

As already explained in Chapter 2 on methodological challenges and Chapter 3 on which confounders were relevant for 'interaction' analysis, key confounders for

both genetic and dietary exposures were included in all analyses. In addition to confounders of the main genetic and dietary exposures, another potential confounder could be related to the interaction effect (i.e. the interaction between the genetic exposure and each covariate as well as the interaction between the environmental exposures and each covariate). This is shown in Equation 6, below, where G= genetic exposure, E= environmental exposure and C= covariates. Keller explained that the interaction term will be biased if i) the covariate is correlated with the genetic variable and the covariate x environment interaction coefficient is non-zero, or ii) the covariate is correlated with the environmental variable and the covariate x gene interaction coefficient is non-zero.[128] Taking a pragmatic approach to avoid the potential overfitting of the primary models and loss of power, the method described by Keller and colleagues was planned as a sensitivity analysis. Therefore, although the possibility of residual confounding from unmeasured confounders remains, the current body of work has gone beyond what previous studies have done to minimise the impact of confounding on the results.

$$f(T2DM) = \beta_o + \beta_G G + \beta_E E + \beta_c C + \beta_{GxE} GE + \beta_{Gxc} GC + \beta_{ExC} EC \qquad (6)$$

### *8.3.4 Type of interaction scale*

In Chapter 3, the difference between the multiplicative and additive interaction scales have been explained. Several studies demonstrate that interactions are scale-dependent.[139,140] Whilst some endorse testing and reporting all interactions on both scales,[141] since additive interactions are only relevant when an interaction has public health relevance, a hypothesis-driven approach was adopted in deciding which interaction scale to use and when. In the replication analysis (Chapter 4), the scale was based on what was applied in the published report. In analyses between GRS and macronutrient intake (Chapter 5), a multiplicative scale was deemed most appropriate given the aim of those analyses was to understand aetiology of T2D, whereas for the interaction analyses between GRS and foods and beverages (Chapter 6), both multiplicative and additive interactions were tested because of the foreseeable public health potential of any

identified interactions. The reason for using multiplicative interactions for the GEWIS was similar to that taken for GRS and macronutrients.

### *8.3.5 Generalisability*

Given that all the study cohorts examined within this PhD were populations of European descent, the question about whether gene-diet interactions exist in other ethnic groups remain unanswered. For example, worldwide, China currently has the highest prevalence of diabetes (11%)[12] and their post-globalisation dietary changes represent a potentially insightful population to examine how transitional patterns of eating (e.g. from traditional Chinese to Western dietary patterns) contribute to rising T2D, under genetic susceptibility. Indeed, some gene-macronutrient interactions have been noted in those of Chinese[192] and Indian descent.[271] Moreover, most of the analyses from this PhD (Chapters 4-7) were conducted within one study population (EPIC-InterAct). This limits the generalisability of findings beyond European populations but also non-EPIC-InterAct European populations. In genetic epidemiology, the practice of *in silico* replication and meta-analyses of several cohorts is more common than for studies of gene-diet interactions, with barriers such as heterogeneity in dietary assessment discussed in Chapter 4. However, EPIC-InterAct is the most suitable study to evaluate the research questions set for this PhD because of the large number of incident cases of T2D and detailed genetic and dietary data available. As well as this, having eight countries represented in EPIC-InterAct allowed us to some extent generalise our findings across Europe. The harmonised assessment and analyses within EPIC-InterAct also minimised possible heterogeneity and therefore differences in potential bias between countries.

Nevertheless, having undertaken most of the analyses in EPIC-InterAct is a limitation of the current findings. Although enlarging the analyses with other similar studies via a meta-analysis (e.g. CHARGE consortium efforts) may be an option to address limitations of generalisability and can improve statistical power, current approaches like this do not fully address long-standing concerns about heterogeneity relating to dietary assessment (e.g. one study uses FFQ whilst another uses 24 hour recall) and analysis (e.g. absence of covariates in certain studies and possible misclassification if variable definitions differ) between pooled studies.[118] In section 8.7, a discussion is made about how future research

should determine what constitutes the most valuable form of replication for gene-diet interactions, whilst considering inherent issues relating to heterogeneity between studies and the impact of important subgroup differences where they exist. UK Biobank is another cohort that may help with evaluating generalisability of interactions in other ethnic groups.

Another limitation of this work is the scope of interactions that could be examined. Whilst the choice in macronutrients was those that have been most commonly consumed and researched, we cannot preclude interactions that may exist between alternative macronutrient (e.g. monosaccharides, 'free sugars') and/or genetic exposures. Therefore, for all of the individual projects conducted within this PhD, broad generalised conclusions have been avoided and this limitation has been highlighted in the respective discussion sections.

### *8.4 Assessment of causality*

The main aim of this PhD was to investigate the existence of and then evaluate the validity and reliability of gene-diet interactions in the aetiology T2D. Given the nature of the findings (i.e. no interactions were identified), next is a brief discussion about how causality of interactions would have been evaluated should any have been identified.

In observational epidemiology, where 'associations' are examined, the 'noise' among free-living populations (i.e. chance, bias, confounding) means that care is needed when making causal inferences. Even when there is internal validity of the findings observed, how it fits within the context of other sources of evidence and its biological plausibility needs to be considered to infer 'causality.' There are three tools/criteria that I would use to aid this assessment, namely the Bradford Hill Criteria, Directed Acyclic Graphs (DAG) and Rothman's sufficient-component-cause model. In 1965, Sir Austin Bradford-Hill proposed a series of criteria for assessing causality in observation research, which was motivated by his attempt to understand the role of smoking in the aetiology of lung cancer.[272] This includes consideration of the strength of association, specificity, dose-response, consistency, temporality, biological plausibility, analogy, coherence and experimental evidence. Although widely used in medical epidemiology, a key criticism of this approach is that it considers a single exposure at a time, which

can be at odds for evaluating conditions of multifactorial nature (e.g. T2D), and for determinants of health that are distant from the individual's biology, such as education.[165,273]

In developing a framework for assessing causality of gene-environment interactions, Geneletti al., suggested combining the Bradford-Hill criteria with a DAG.[165] They advised DAGs to help visualise the complexity of the interaction and Hill's criteria to determine which edges can be considered causal. Alternatively, Rothman in 1976 proposed the 'sufficient-component-cause model' or 'causal pie.'[134] This overcame the above problems, as it attributes disease development to multiple causes, each called a 'component cause' and only a complete pie is considered a causal pathway, called a 'sufficient cause.' Unlike the Bradford-Hill criteria, it does not reduce disease development to a single 'component cause' but recognises that multifactorial conditions may arise from an accumulation of variable factors that could potentially differ between people but must be present simultaneously to cause the disease. This model resonates with research on gene-diet interactions, at least with one hypothesis about the synergistic impact of the presence of both genetic and dietary risk factors in the development of T2D.

Specifically for Hill's criterion of consistency, the findings from this PhD demonstrated internal consistency for no interaction based on various methods used, including weighted and unweighted GRS, combined GRS and single SNPs and in some instances on both multiplicative and additive interaction scales. However, although replication would strengthen the reliability of either a null or positive interaction finding, whether the dynamic nature of interactions lends to be replicated or consistent across what may appear to be demographically similar populations, is questionable. This will be further discussed in section 8.7 on implications.

Rothman's sufficient-component-cause model helps to explain the dynamic nature of interactions and offers an alternative explanation for the lack of interactions identified within this PhD. It is well accepted that whilst some dietary effects persist across the lifespan, some are transitory. This is evident with nutritional requirements that adapt to critical periods associated with growth and

development, including from infancy to adolescence and during pregnancy. Several studies have identified exposure modification by age, including the differential effect of cigarette smoking on cancer risk in mice at different developmental stages,[274] and that high-fat diets consumed by maternal mice had a greater impact on cancers in the offspring than high-fat diets consumed by the offspring postnatally.[275] Genetic expression also changes over time, observed at the day-to-day level, reflecting constantly changing environmental exposures.[276] Indeed, higher order multi-way interactions (e.g. sex-specific effects for the interaction between *IRS1* and macronutrient intake)[195,277,278] have been suggested and these pieces of evidence together suggest that gene-environment interactions may occur only during critical periods of time and/or are sensitive to other factors such as sex and context. Therefore, it may only be possible to detect certain interactions in certain populations at certain times. Rothman's model may help to explain why the absence of gene-diet interactions in this PhD does not necessarily preclude the existence of these interactions, but rather that the current context in which this particular population is studied may be insufficient to bring about a detectable interaction.[222] In other words, either another modifier (e.g. sex) is present at a higher degree and may mask the gene-diet interactions of interest or that only under certain conditions (presence of another modifier) will the interactions of interest become apparent. However, methodological limitations such as sample size restrictions did not enable this to be tested in the current PhD.

Certainly, several fields including cancer epidemiology and psychiatry are actively researching gene-environment interactions, with several theories supporting its existence including the diathesis-stress model, differential susceptibility, plasticity alleles, biopsychosocial model and evolutionary developmental psychology.[202,279] Across all these fields, similar methodological challenges exist in the context of the aforementioned limitations of observational studies. Therefore, from a fundamental standpoint, interactions between genes and environmental factors including dietary, must contribute to the development of multifactorial phenotypes (including diseases). None were identified between the specific exposures that were examined in this PhD, for the risk of developing T2D in this specific population. Therefore, before causality can be examined, it may be

more valuable to evaluate the suitability of the current methods and study designs used to examine interactions. And if interactions are identified the clinical significance of the magnitude of interaction effect will determine its applicability.

### *8.5 Clinical utility of genetic risk information*

Apart from the validity of gene-diet interactions, a commonly asked question is whether disclosure of genetic risk information would lead to behaviour change and subsequent improvement in health outcomes (clinical utility). Certainly, genetic risk may be a potent motivator for behaviour change because of its biological accuracy and personal salience, which is consistent with the Health Belief Model.[280] Early research indeed indicated that the provision of personalised genetic information favourably influenced screening behaviours and medication adherence for individuals at risk of familial cancers, often involving Mendelian inheritance with high penetrance genetic variants.[281] However, this cannot be assumed for the adoption of more complex 'lifestyle' health-related behaviours, such as dietary modification, that are required to be adopted and sustained in order to reduce the risk of developing cardiometabolic disorders such as obesity, T2D and cardiovascular disease (CVD).

As an adjunct to this PhD, I led on a systematic review and meta-analysis of RCTs undertaken in the context of cardiometabolic disorders (obesity, T2D, CVD) to investigate: 1) the effect of genetic risk testing and communication on perceived and actual motivation to engage in risk reduction lifestyle modification (diet and physical activity); 2) the effect of genetic risk testing and communication on actual lifestyle modification and 3) clinical outcomes.[282] From 13 RCTs (N=8,426 participants) there were no consistent effect of genetic risk on actual motivation for weight loss, perceived motivation for dietary change (control compared to genetic risk group standardised mean difference (SMD) −0·15; 95% CI −1·03, 0·73, P=0·74) or actual change in dietary behaviour. Similar results were observed for actual weight loss (control compared to high genetic risk SMD 0·29 kg; 95% CI −0·74, 1·31, P=0·58). This review found no clear or consistent evidence that genetic risk communication alone either raises motivation or translates into actual change in dietary intake or physical activity to reduce the risk of cardiometabolic disorders in adults.[282]

Our findings were consistent with the updated Cochrane Systematic Review for dietary (SMD: 0.12, 95%CI: -0.00 to 0.24, p:0.05) and physical activity behaviours (SMD: -0.03, 95%CI: -0.14 to 0.07, p:0.54).[283] The largest European-wide internet based RCT on personalised nutrition (Food4Me, N=1,269) also did not observe any difference in dietary intake between groups who had received personalised dietary advice according to participant's current dietary intake (L1), phenotypic (weight, BMI, waist circumference and blood markers) (L2) or genotypic information (L3), after 6 months (e.g. p value for the difference between Healthy Eating Index [HEI] in those in L1 compared to those in L2+L3= 0.693).[284] Nevertheless, it did demonstrate that simple personalisation, including by information on dietary intake, was more effective in promoting healthy eating than conventional dietary advice.[284] This suggests that the process of personalisation (i.e. tailoring advice) rather than the tool used to convey personalisation (e.g. genes) appears more important in promoting healthy eating. To ensure public welfare in engaging with genetic susceptibility information, policy makers need to enforce stricter regulation of direct to consumer services offering personalised nutrition by genotype.

### *8.6 Assessing the causality of diet and disease using genetics*

A related area of interest is the application of genetics in nutritional epidemiology, which is helping us to understand the aetiology of cardiometabolic conditions, including T2D. Mendelian Randomisation (MR) is a statistical method that enables inferences to be made about causal relationships within observational studies.[285] Genes associated with a particular exposure are used as genetic proxies for the lifetime environmental exposure of interest without the complication of reverse causality and confounding, which are known limitations of observational nutrition studies.[286]

I contributed to a meta-analysis of GWAS studies (as part of the CHARGE Consortium, unpublished) that investigated genetic variants associated with percentage of energy from protein, fat or carbohydrate intake. The associated genetic variants were used to develop genetic instruments to examine the respective causal relationship between these macronutrient intake exposures and BMI, a strong risk factor for T2D. There was reasonably clear evidence to support

causality for higher genetically predicted BMI influencing higher genetically predicted protein intake (Table 9-1). A previous report highlighted that the *FTO* locus, a BMI associated locus, was associated with higher protein intake.[287] This is a finding we now extend to other BMI-raising alleles suggesting that higher BMI is associated with higher reported protein intake and not specific to *FTO*. An alternative explanation of the findings may be that there is a shared biological 'soil' between macronutrients and BMI. Given that most BMI SNPs act through the brain and may be involved in energy homeostasis,[28] perhaps these MR findings indicate BMI driving cognitive intake preferences possibly via increasing TEI, which protein intake may be a marker for. Secondly, all three genetically predicted macronutrients were associated with genetically predicted BMI (Table 8-1). This is because the genetic variant at *FGF21* was included in the genetic instrument for all three macronutrients given its shared association. Whether this reflects substitution of macronutrients as a proportion of TEI is difficult to assess, so, causality between any one specific macronutrient group with BMI cannot be determined.

**Table 7-1: Mendelian Randomisation results for the relationship between genetically predicted protein intake and genetically predicted BMI**

| Genetically predicted exposure | Genetically predicted outcome (IVW MR) | Effect estimates |
|---|---|---|
| ↑ 1 standard deviation in BMI due to a 94-SNP polygenic risk score (excluding FTO) | ↑0.58%TEI from protein intake | *SE* =0.08, $P_{ivw}$ =9.88E-13 |
| ↑ 1% of TEI from protein intake | ↑0.09 kg/m$^2$ (BMI) | *SE* =0.03, $P_{ivw}$ =6.92E-04 |
| ↑ 1% of TEI from fat intake | ↑0.05 kg/m$^2$ (BMI) | *SE* =0.02, $P_{ivw}$ =6.92E-04 |
| ↑ 1% of TEI from carbohydrate intake | ↓0.06 kg/m$^2$ (BMI) | *SE* =0.01, $P_{ivw}$ =1.12E-09 |

Note: 97 SNPs were associated with BMI (Locke et al., 2015) but we included 94 SNPs because 2 had no proxies and another genetic variant in FTO locus was excluded.
Abbreviation: TEI- total energy intake, MR- Mendelian randomisation, IVW- inverse variance weighted

## *8.7 Implications and recommendations*

In summary, the findings from this PhD indicate that none of the gene-diet interactions examined within this PhD significantly contribute to the development of T2D. During this PhD, the inherent limitations to the research methods currently adopted for investigating gene-diet interactions is clear. Therefore, improvement in methods is a key priority going forward for this field. This and the findings from this PhD lead to several important implications for

future research to better understand if gene-diet interactions are truly important in the aetiology of T2D.

Firstly, in conducting the systematic review, where the available literature on gene-macronutrient interactions and T2D was critiqued, studies were highly heterogeneous and this have been previously highlighted as a barrier to meaningful quantitative pooling of data.[154] There has been on-going criticism about the quality of studies available (from the review in Chapter 4 and many others).[105,173,202] Alongside this, in several projects within this PhD, methodological decisions were based on judgement using available knowledge rather than previously evaluated best practice such as treating macronutrient exposures and the interaction term as linear variables. These issues make it difficult to derive clear conclusions about gene-diet interactions. Moreover, currently there are no published standards in conducting gene-environment interactions, unlike the STREGA statement used for evaluating GWAS.[143] Despite on-going methodological developments in this field, there needs to be a formally agreed upon standard in conducting gene-environment interactions that is accepted by the scientific community, to guide more rigorous future research and assist more informed evaluation and interpretation of published interaction studies. Efforts are emerging, with a proposed set of guidelines for cancer epidemiology,[144] so how well this can be applied to other conditions such as diabetes will be of interest. I have also provided some suggestions towards this goal, from our own work (Appendix A).

Secondly, it is clear that in observational nutritional epidemiology examining macronutrient intake, associations with disease requires isocaloric macronutrient substitution modelling for accuracy and findings to be interpretable. This has been neglected within studies of gene-macronutrient interactions, at least within the cardiometabolic literature. Although findings from this PhD which compare analyses with and without isocaloric macronutrient interactions are not materially different (Chapter 4 and 5), it is methodologically more accurate to apply isocaloric macronutrient substitution. Therefore, it is the hope that this work will provide the impetus for future gene-macronutrient interaction research to adopt this practice.

Thirdly, even with all the challenges of conducting replication, mentioned in Chapter 4 including difficulties in finding the appropriate replication population, study design, exposure assessment and modelling of covariates to name a few, replication adds to the body of evidence necessary to evaluate gene-diet interactions and is an important process. By minimising the possibility of chance findings (Winner's curse) this aims to improve the credibility of findings. Although the best form of 'replication,' is yet to be defined, some examples of approaches that may address the issue of heterogeneity are presented next. Recently, a study using UK Biobank to examine the interaction between *FTO* and lifestyle factors on the risk of obesity cleverly divided their population into three samples.[107] Although intended to better control for population substructure, the consistency in methods enabled appropriate replication and consistency in findings between those who were classified as 'British' and those more ethnically diverse. Together, this strengthened the inference of detected interactions. In cases where opportunities to access such large and well-phenotyped databases may not be possible, suggestions have been made for accompanying interactions with findings using complementary methods (e.g. functional studies in vitro) to aid interpretation.

An alternative method may be to conduct individual participant data (IPD) meta-analysis that enables harmonisation of terminology, analysis method and researcher interpretation using raw participant data (one- stage approach). Tools such as DataSHaPER and guidelines such as that written by Maelstrom Research have been developed to facilitate retrospective harmonisation between studies, [288,289] however this process is resource-intensive. Hence, once platforms such as InterConnect (http://www.interconnect-diabetes.eu/) are established, with the aim of investigating gene-environment interactions in mind, it will be a valuable resource for researchers. Studies from the Emerging Risk Factors Collaboration adopt an IDP approach to examine lipid, inflammatory and other markers in the development of cardiovascular disease, also offer existing examples (http://www.phpc.cam.ac.uk/ceu/erfc/erfc-publications/). Therefore, investing time and research into deciding what will be the most valid and feasible method/s to constitute a 'replication,' will be a valuable pursuit that will resolve this problem for the interaction field.

Although in this PhD no convincing interactions have been identified, I have learnt that a rigorously conducted study with null findings declares more influence than a positive finding from a less well-conducted study. I agree with others that future interaction studies should start off with the expectation to disprove their interactions and only when it is robust to a battery of tests, like replication or other, should that interaction be confidently accepted. Despite what may be considered a conservative approach, this may be necessary to reinstate confidence back into research conducted within this field.

Fourthly, a recurring theme throughout this thesis is the difficulty in accounting for the dynamics of gene-diet interactions (e.g. presence of other factors such as age, sex, etc), as explained in section 8.4 on causality. Therefore, this means that the existence of gene-diet interaction cannot be fully invalidated even if none were identified in our case within a particular population at a particular point in time. This was explained by how higher order multi-way interactions exist and that current epidemiological methods are limited by their sample size and issues such as confounding to detect this. Suggestions have been made about how to study this, ranging from using long-term follow-up studies with repeated dietary assessment, systems biology, experimental studies under strict regulation of other environment exposures, and recall by genotype dietary interventions.[96,120] Carvalho-Wells and colleagues used the recall by genotype approach when investigating ApoE and fat intake on plasma triglyceride and CRP levels, to minimise 'noise' and improve statistical power by recruiting participants according to their ApoE genotype.[115] Additionally, whilst the method of MR has predominantly been used to assess the causal role of a single exposure with a single outcome, it may help in determining the causality of gene-diet interactions with T2D.[290] To my knowledge, no studies have applied the MR framework in examining gene-diet interactions and may be worthwhile to investigate. All aforementioned approaches are beyond the scope of present thesis and it is up to future studies to investigate how best to do this.

The GEWIS of macronutrient intake and T2D remain on-going and future plans for research have already been described in Chapter 7.

Lastly, these findings are important for health professionals who are involved in nutrition practice (e.g. clinicians and dietitians) to recognise that no change to current clinical or public health advice regarding macronutrient intake and T2D prevention, is warranted. Present evidence from gene-diet interaction research and studies investigating the clinical utility of genetic risk communication on behaviour change gives health professionals evidence not to support the use of genetically 'personalised nutrition' services that are currently commercially available. Nutrition practitioners are responsible for correcting misinformation, where necessary, so researchers should strive to reach this target audience. For example, I have written articles for an international resource database for dietitians, called Practice-based Evidence in Nutrition (http://www.pennutrition.com/index.aspx). However, health professionals can only do so much as to advise the public, it is therefore under the power of policymakers to instate stricter regulations on commercial enterprises to protect public safety and welfare. Although a balance between regulation and freedom for creativity is needed not to stifle meaningful scientific pursuits in this field, honest and realistic communication about the current and future scientific potential of this research is necessary.

### *8.8 Conclusion*

The totality of the evidence produced from original research during this PhD has meaningfully added to the available literature on gene-diet interactions and T2D, to suggest that the particular interactions examined within this PhD do not contribute to the aetiology of T2D. Despite the benefits of personalisation in nutritional therapy, at present on the basis of the lack of empirical evidence for robustly validated gene-diet interactions, currently there is no evidence to support genetic personalisation of dietary advice to prevent T2D. This body of research represents the most comprehensive and systematic investigation of gene-macronutrient interactions for the risk of developing T2D to date, and one I thoroughly enjoyed contributing to. Considerable methodological limitations still exist in this research field, highlighting the urgent need for re-evaluating how

interactions should be studied and validated before definitive conclusions can be made about their independent contribution to T2D. In keeping with the view proposed by Lancelot Hogben relating to the interdependence of Nature and Nurture, which underpins the study of gene-environment interactions, it may indeed be time to develop novel methods that also stretch beyond the summing of its component parts (namely, genetics and nutrition).

# *List of projects*

| Projects that I led on |
|---|
| 1. Interaction between genetic variants and macronutrient intake on the risk of developing type 2 diabetes: systematic review and findings from EPIC- InterAct |
| 2. Interplay between genetic predisposition, macronutrient intake and Type 2 Diabetes incidence. Analysis across 8 European countries: EPIC-InterAct Study |
| 3. Interplay between genetic predisposition, food and beverage intake for Type 2 Diabetes incidence. Analysis across 8 European countries: EPIC-InterAct Study |
| 4. Genome-wide-interaction study of macronutrient intake and incident Type 2 Diabetes: EPIC-InterAct Study |
| 5. The effect of communicating the genetic risk of cardiometabolic disorders on motivation and actual engagement in preventative lifestyle modification and clinical outcome: a systematic review and meta-analysis of randomised controlled trials |
| **Projects that I have contributed to and have been discussed within this thesis** |
| 1. Macronutrient intake genome-wide-association-study (CHARGE Consortium) |
| **Projects that I have contributed to but were not discussed within this thesis** |
| 1. Circulating saturated fatty acid genome-wide-association-study (EPIC-InterAct) |
| 2. Genome-wide-interaction study of circulating polyunsaturated fatty acids and C-Reactive-Protein (CHARGE Consortium) |
| 3. Genetic risk score and dietary pattern interaction analysis for incident Type 2 Diabetes: EPIC-InterAct |
| 4. Association between circulating vitamin D and Type 2 Diabetes: EPIC-InterAct |
| 5. Circulating fatty acids and trends in EPIC-Norfolk |

# *References*


1. Hogben LT. Genetic Principles in Medicine and Social Science. London: Williams & Norgate; 1932.
2. Tuomi T, Santoro N, Caprio S, Cai M, Weng J, Groop L. The many faces of diabetes: a disease with increasing heterogeneity. Lancet 2014;383(9922):1084–94.
3. Florez JC. Leveraging Genetics to Advance Type 2 Diabetes Prevention Identifying Individuals at Increased Risk. PLoS Med 2016;1–7.
4. Hindy G, Sonestedt E, Ericson U, Jing X-J, Zhou Y, Hansson O, et al. Role of *TCF7L2* risk variant and dietary fibre intake on incident type 2 diabetes. Diabetologia 2012;55(10):2646–54.
5. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè A V, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44(9):981–90.
6. WHO. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Geneva: 1999.
7. ADA. Diagnosis and classification of diabetes mellitus. Diabetes Care 2014;37 Suppl 1(January):S81-90.
8. WHO, IDF. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. Geneva: 2006.
9. WHO. Use of Glycated Haemoglobin ( HbA1c ) in the Diagnosis of Diabetes Mellitus. Geneva: 2011.
10. ADA. Standards of Medical Care in Diabetes - 2017. Diabetes Care 2017;40 (Suppl.:S44-47).
11. NCD-RisC. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4·4 million participants. Lancet 2016;387:1513–30.
12. IDF. International Diabetes Federation. IDF Diabetes Atlas, 7th edn. Belgium: 2015.
13. Fowler MJ. Microvascular and Macrovascular Complications of Diabetes. Clin Diabetes 2014;29(3):116–22.
14. Franks PW, Mccarthy MI. Exposing the exposures responsible for type 2 diabetes and obesity. Science 2016;354(6308):69–73.
15. Bommer C, Heesemann E, Sagalova V, Manne-Goehler J, Atun R, Bärnighausen T, et al. The global economic burden of diabetes in adults aged 20–79 years: a cost-of-illness study. Lancet Diabetes Endocrinol 2017;8587(17):1–8.

16. World Health Organization. Global Report on Diabetes. 2016.

17. Kahn SE, Cooper ME, Del Prato S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. Lancet 2014;383(9922):1068–83.

18. Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: principles of pathogenesis and therapy. Lancet 2010;365(9467):1333–46.

19. Mozaffarian D. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity: A Comprehensive Review. Circulation 2016;133:187–225.

20. Bell J a, Kivimaki M, Hamer M. Metabolically healthy obesity and risk of incident type 2 diabetes: a meta-analysis of prospective cohort studies. Obes Rev 2014;15(6):504–15.

21. White M. Population Approaches to Prevention of Type 2 Diabetes. PLoS Med 2016;13(7):10–3.

22. Conklin AI, Monsivais P, Khaw K-T, Wareham NJ, Forouhi NG. Dietary diversity, diet cost, and incidence of type 2 diabetes in the United Kingdom: a prospective cohort study. PLoS Med 2016;13(7):e1002085.

23. Stokols D. Translating Social Ecological Theory into guidelines for community health promotion. Am J Heal Promot 1996;10(4):282–98.

24. NICE. National Institute for Health and Care Excellence. Preventing type 2 diabetes overview [Internet]. Flowchart2017 [cited 2017 May 4];Available from: https://pathways.nice.org.uk/pathways/preventing-type-2-diabetes

25. Almgren P, Lehtovirta M, Isomaa B, Sarelin L, Taskinen MR, Lyssenko V, et al. Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. Diabetologia 2011;54(11):2811–9.

26. Willemsen G, Ward KJ, Bell CG, Christensen K, Bowden J, Dalgård C, et al. The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. Twin Res Hum Genet 2015;18(6):762–71.

27. Mahajan A, Go M, Zheng W, Below J, Gaulton K, Ferreira T. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 2014;46(3):234–44.

28. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518:197–206.

29. Scott R a, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet 2012;44(9):991–1005.

30. Lotta L, Gulati P, Day FR, Payne F, Ongen H, van de Bunt M, et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. Nat Genet 2016;49(1):17–26.

31. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47(11):1236–41.

32. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 2006;38(3):320–3.

33. Wang J, Kuusisto J, Vänttinen M, Kuulasmaa T, Lindström J, Tuomilehto J, et al. Variants of transcription factor 7-like 2 (TCF7L2) gene predict conversion to type 2 diabetes in the Finnish Diabetes Prevention Study and are associated with impaired glucose regulation

and impaired insulin secretion. Diabetologia 2007;50(6):1192–200.

34. Fuchsberger C, Flannick J, Teslovich T, Mahajan A, Agarwala V, Gaulton K. The genetic architecture of type 2 diabetes. Nature 2016;536(7614):41–7.

35. Thomas D. Gene-environment-wide association studies: emerging approaches. Nat Rev Genet 2010;11(4):259–72.

36. Prevention D. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. Lancet Diabetes Endocrinol 2015;3(11):866–75.

37. Li G, Zhang P, Wang J, An Y, Gong Q, Gregg EW, et al. Cardiovascular mortality , all-cause mortality, and diabetes incidence after lifestyle intervention for people with impaired glucose tolerance in the Da Qing Diabetes Prevention Study: a 23-year follow-up study. Lancet Diabetes Endocrinol 2014;8587(14):1–9.

38. Lindström J, Peltonen M, Eriksson JG, Ilanne-Parikka P, Aunola S, Keinänen-Kiukaanniemi S, et al. Improved lifestyle and decreased diabetes risk over 13 years: Long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). Diabetologia 2013;56(2):284–93.

39. Knowler W, Barrett-Connor E, Flowler S, Hamman R, Lachin J, Walker E, et al. Reduction in the incidence of type 2 diabetess with lifestyle intervention or metformin. N Engl J Med 2002;346(6):393–403.

40. WHO/FAO. Diet, nutrition and the prevention of chronic diseases. Report of a joint WHO/FAO Expert Consultation. 2003.

41. Dyson P a, Kelly T, Deakin T, Duncan a, Frost G, Harrison Z, et al. Diabetes UK evidence-based nutrition guidelines for the prevention and management of diabetes. Diabet Med 2011;28(11):1282–8.

42. Ley SH, Hamdy O, Mohan V, Hu FB. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. Lancet 2014;383(9933):1999–2007.

43. Martinez-Gonzalez MA, Salas-Salvado J, Estruch R, Corella D, Fito M, Ros E. Benefits of the Mediterranean Diet: Insights From the PREDIMED Study. Prog Cardiovasc Dis 2015;58(1):50–60.

44. Cooper AJ, Forouhi NG, Ye Z. Fruit and vegetable intake and type 2 diabetes : EPIC-InterAct prospective study and meta-analysis. Eur J Clin Nutr 2012;66(10):1082–92.

45. Aune D, Norat T, Romundstad P, Vatten LJ. Whole grain and refined grain consumption and the risk of type 2 diabetes: A systematic review and dose-response meta-analysis of cohort studies. Eur J Epidemiol 2013;28(11):845–58.

46. Afshin A, Micha R, Khatibzadeh S, Mozaffarian D. Consumption of nuts and legumes and risk of incident ischemic heart disease , stroke , and diabetes : a systematic review and. Am J Clin Nutr 2014;100:278–89.

47. Gijsbers L, Ding EL, Malik VS, Goede J De, Geleijnse JM, Soedamah-muthu SS. Consumption of dairy foods and diabetes incidence : a dose-response meta-analysis of observational studies. Am J Clin Nutr 2016;103:1111–24.

48. Aune D, Orat T, Romundstad P, Vatten LJ. Dairy products and the risk of type 2 diabetes : a systematic review and dose-response meta-analysis of cohort studies. Am J Clin Nutr 2013;98:1066–83.

49. Ding M, Bhupathiraju SN, Chen M, Van Dam RM, Hu FB. Caffeinated and decaffeinated coffee consumption and risk of type 2 diabetes: A systematicreview and a dose-response meta-analysis. Diabetes Care 2014;37(2):569–86.

50. Yang J, Mao Q-X, Xu H-X, Ma X, Zeng C-Y. Tea consumption and risk of type 2 diabetes mellitus: a systematic review and meta-analysis update. BMJ Open 2014;4(7):e005632.

51. Wallin A, Di Giuseppe D, Orsini N, Patel PS, Forouhi NG, Wolk A. Fish consumption, dietary long-chain n-3 fatty acids, and risk of type 2 diabetes: systematic review and meta-analysis of prospective studies. Diabetes Care 2012;35(4):918–29.

52. Tamze M, Virtanen J, Gaziano JM. Egg consumption and risk of type 2 diabetes: a meta-analysis of prospective studies. Br J Nutr 2016;103(2):474–80.

53. Imamura F, O'Connor L, Ye Z, Mursu J, Hayashino Y, Bhupathiraju SN, et al. Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. BMJ 2015;351:1–12.

54. Hu E a, Pan A, Malik V, Sun Q. White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. BMJ 2012;344(March):e1454.

55. Pan A, Sun Q, Bernstein AM, Schulze MB, Manson JE, Willett WC, et al. Red meat consumption and risk of type 2 diabetes : 3 cohorts of US adults and an updated meta-analysis. Am J Clin Nutr 2011;194:1088–96.

56. Jacobs DR, Tapsell LC. Food, not nutrients, is the fundamental unit in nutrition. Nutr Rev 2007;65(10):439–50.

57. Carlberg C, Campbell MJ. Vitamin D receptor signaling mechanisms: Integrated actions of a well-defined transcription factor. Steroids 2013;78(2):127–36.

58. Johnston BC, Kanters S, Bandayrel K, Wu P, Naji F, Siemieniuk RA., et al. Comparison of Weight Loss Among Named Diet Programs in Overweight and Obese Adults. JAMA 2014;312(9):923–33.

59. Lindström J, Peltonen M, Eriksson JG, Louheranta A, Fogelholm M, Uusitupa M, et al. High-fibre, low-fat diet predicts long-term weight loss and decreased type 2 diabetes risk: the Finnish Diabetes Prevention Study. Diabetologia 2006;49(5):912–20.

60. Martinez JA, Navas-Carretero S, Saris WHM, Astrup A. Personalized weight loss strategies-the role of macronutrient distribution. Nat Rev Endocrinol 2014;10(12):749–60.

61. Hooper L, Abdelhamid A, Moore HJ, Douthwaite W, Skeaff CM, Summerbell CD. Effect of reducing total fat intake on body weight: systematic review and meta-analysis of randomised controlled trials and cohort studies. BMJ 2012;345:1–15.

62. Tobias DK, Chen M, Manson JAE, Ludwig DS, Willett W, Hu FB. Effect of low-fat diet interventions versus other diet interventions on long-term weight change in adults: A systematic review and meta-analysis. Lancet Diabetes Endocrinol 2015;3(12):968–79.

63. Alhazmi A, Stojanovski E, McEvoy M, Garg ML. Macronutrient Intakes and Development of Type 2 Diabetes: A Systematic Review and Meta-Analysis of Cohort Studies. J Am Coll Nutr 2012;31(4):243–58.

64. Mansoor N, Vinknes KJ, Veierød MB, Retterstøl K. Effects of low-carbohydrate diets v. low-fat diets on body weight and cardiovascular risk factors: a meta-analysis of randomised controlled trials. Br J Nutr 2016;115:466–79.

65. Hu T, Mills KT, Yao L, Demanelis K, Eloustaz M, Yancy WS, et al. Systematic Reviews and Meta- and Pooled Analyses Effects of Low-Carbohydrate Diets Versus Low-Fat Diets on Metabolic Risk Factors : A Meta-Analysis of Randomized Controlled Clinical Trials. Am J Epidemiol 2012;176(7):S45-54.

66. South Beach Diet [Internet]. 2017;Available from: http://www.southbeachdiet.com/diet/dr-agatston/index.jsp

67. Real Meal Revolution: Home of Banting [Internet]. 2017;Available from: https://realmealrevolution.com/

68. Naude CE, Schoonees A, Senekal M, Young T, Garner P, Volmink J. Low Carbohydrate versus Isoenergetic Balanced Diets for Reducing Weight and Cardiovascular Risk: A Systematic Review and Meta-Analysis. PLoS One 2014;9(7):e100652.

69. Shai I, Schwarzfuchs D, Henkinm Y, Shahar D, Witkow S, Greenberg I. Weight Loss with a Low-Carbohydrate, Mediterranean, or Low-Fat Diet. N Engl J Med 2008;359(3):229–41.

70. Schulze MB, Schulz M, Heidemann C, Schienkiewitz A, Hoffmann K, Boeing H. Carbohydrate intake and incidence of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Br J Nutr 2008;99(5):1107–16.

71. Rietman A, Schwarz J, Tomé D, Kok FJ, Mensink M. High dietary protein intake, reducing or eliciting insulin resistance? Eur J Clin Nutr 2014;68(9):973–9.

72. Santesso N, Akl E a, Bianchi M, Mente a, Mustafa R, Heels-Ansdell D, et al. Effects of higher- versus lower-protein diets on health outcomes: a systematic review and meta-analysis. Eur J Clin Nutr 2012;66(7):780–8.

73. Wycherley TP, Moran LJ, Clifton PM, Noakes M, Brinkworth GD. Effects of energy-restricted high-protein, low-fat compared with standard-protein, low-fat diets: a meta-analysis of randomized controlled trials 1–3. Am J Clin Nutr 2012;96(2):1281–98.

74. Schwingshackl L, Hoffmann G. Long-term effects of low-fat diets either low or high in protein on cardiovascular and metabolic risk factors: a systematic review and meta-analysis. Nutr J 2013;12:48.

75. Shang X, Scott D, Hodge AM, English DR, Giles GG, Ebeling PR, et al. Dietary protein intake and risk of type 2 diabetes: results from the Melbourne Collaborative Cohort Study and a meta-analysis of prospective studies. Am J Clin Nutr 2016;104(5):1352–65.

76. Sluijs I, Beulens JWJ, van der A DL, Spijkerman AMW, Grobbee DE, van der Schouw YT. Dietary intake of total, animal, and vegetable protein and risk of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)-NL study. Diabetes Care 2010;33(1):43–8.

77. Ericson U, Sonestedt E, Gullberg B, Hellstrand S, Hindy G, Wirfält E, et al. High intakes of protein and processed meat associate with increased incidence of type 2 diabetes. Br J Nutr 2013;109(6):1143–53.

78. Imamura F, Micha R, Wu JHY, de Oliveira Otto MC, Otite FO, Abioye AI, et al. Effects of Saturated Fat, Polyunsaturated Fat, Monounsaturated Fat, and Carbohydrate on Glucose-Insulin Homeostasis: A Systematic Review and Meta-analysis of Randomised Controlled Feeding Trials. PLoS Med 2016;13(7):1–18.

79. Chowdhury R, Warnakula S, Kunutsor S, Crowe F, Ward HA, Johnson L, et al. Association of Dietary , Circulating , and Supplement Fatty Acids With Coronary Risk : A Systematic Review and Meta-analysis. Ann Intern Med 2014;160:398–406.

80. de Souza RJ, Mente A, Maroleanu A, Cozma AI, Ha V, Kishibe T, et al. Intake of saturated and trans unsaturated fatty acids and risk of all cause mortality, cardiovascular disease, and type 2 diabetes: systematic review and meta-analysis of observational studies. BMJ 2015;351:1–16.

81. Mozaffarian D, Micha R, Wallace S. Effects on Coronary Heart Disease of Increasing Polyunsaturated Fat in Place of Saturated Fat: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. PLoS Med 2010;7(3):1–10.

82. Li Y, Hruby A, Bernstein AM, Ley SH, Wang DD, Chiuve SE, et al. Saturated Fats Compared with Unsaturated Fats and Sources of Carbohydrates in Relation to Risk of Coronary Heart Disease A Prospective Cohort Study. J Am Coll Cardiol 2015;66(14):1538–48.

83. Forouhi NG, Imamura F, Sharp SJ, Koulman A, Schulze MB, Zheng J, et al. Association of Plasma Phospholipid n-3 and n-6 Polyunsaturated Fatty Acids with Type 2 Diabetes: The EPIC-InterAct Case-Cohort Study. PLoS Med 2016;13(7):1–17.

84. Viguiliouk E, Stewart SE, Jayalath VH, Ng AP, Mirrahimi A, de Souza RJ, et al. Effect of replacing animal protein with plant protein on glycemic control in diabetes: A systematic review and meta-analysis of randomized controlled trials. Nutrients 2015;7(12):9804–24.

85. Malik VS, Li Y, Tobias DK, Pan A, Hu FB. Dietary Protein Intake and Risk of Type 2 Diabetes in US Men and Women. Am J Epidemiol 2016;183(8):715–28.

86. The InterAct Consortium. Dietary fibre and incidence of type 2 diabetes in eight European countries: the EPIC-InterAct Study and a meta-analysis of prospective studies. Diabetologia 2015;58(7):1394–408.

87. Smith CE, Tucker KL. Health benefits of cereal fibre: a review of clinical trials. Nutr Res Rev 2011;24:118–31.

88. Gibb RD, McRorie JW, Russell DA, Hasselblad V, D'Alessio DA. Psyllium fiber improves glycemic control proportional to loss of glycemic control : a meta-analysis of data in euglycemic subjects , patients at risk of type 2 diabetes mellitus , and patients being treated for type 2 diabetes mellitus. Am J Clin Nutr 2015;102:1604–14.

89. Yao B, Fang H, Xu W, Yan Y, Xu H, Liu Y, et al. Dietary fiber intake and risk of type 2 diabetes: a dose-response analysis of prospective studies. Eur J Epidemiol 2014;29(2):79–88.

90. Rylander R. Letters to the Editor: high protein, low carbohydrate and mineral balance. Am J Clin Nutr 2011;93:1152–6.

91. Calton JB. Prevalence of micronutrient deficiency in popular diet plans. J Int Soc Sports Nutr 2010;7(1):24.

92. Müller M, Kersten S. Nutrigenomics : goals and strategies. Nat Rev Genet 2003;4(April):315–22.

93. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet 2005;6(4):287–98.

94. Williams RA, Mamotte CDS, Burnett JR. Phenylketonuria: an inborn error of phenylalanine metabolism. Clin Biochem Rev 2008;29(1):31–41.

95. Franks PW, Pearson E, Florez JC. Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects. Diabetes Care 2013;36(5):1413–21.

96. Franks PW, Paré G. Putting the Genome in Context: Gene-Environment Interactions in Type 2 Diabetes. Curr Diab Rep 2016;16(7):57.

97. Gibney MJ, Walsh MC. The future direction of personalised nutrition: my diet, my phenotype, my genes. Proc Nutr Soc 2013;72(2):219–25.

98. Ferguson LR, De Caterina R, Görman U, Allayee H, Kohlmeier M, Prasad C, et al. Guide and Position of the International Society of Nutrigenetics/Nutrigenomics on Personalised Nutrition: Part 1 - Fields of Precision Nutrition. J Nutrigenet Nutrigenomics 2016;9:12–27.

99. Martínez JA. Perspectives on personalized nutrition for obesity. J Nutrigenet Nutrigenomics 2014;7(1):1–3.

100. MRC. Stratified Medicine Initiative- call for proposals for disease-focused partnerships to

stratify for patient benefit [Internet]. [cited 2017 Mar 16];Available from: https://www.mrc.ac.uk/funding/browse/smi/stratified-medicine-initiative-disease-focused-partnerships-for-patient-benefits/

101. The White House Factsheet: President Obama's Precision Medicine Initiative [Internet]. [cited 2017 Mar 16];Available from: https://obamawhitehouse.archives.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative

102. Rhea M, Bettles C. Future Changes Driving Dietetics Workforce Supply and Demand: Future Scan 2012-2022. J Acad Nutr Diet 2012;112(3):S10–24.

103. García-Closas M, Malats N, Silverman D, Dosemeci M, Kogevinas M, Hein DW, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. Lancet 2005;366(9486):649–59.

104. Caspi A, Sugden K, Moffitt T, Taylor A, Craig IW, Harrington H, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science 2003;301(5631):386–9.

105. Duncan LE, Keller MC. A Critical Review of the First 10 Years of Candidate Gene-by - Environment Interaction Research in Psychiatry. Am J Psychiatry 2011;168:1041–9.

106. Kilpeläinen TO, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E, et al. Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. PLoS Med 2011;8(11):e1001116.

107. Young AI, Wauthier F, Donnelly P. Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. Nat Commun 2016;7(12724):1–12.

108. Celis-Morales C, Marsaux CFM, Livingstone KM, Navas-Carretero S, San-Cristobal R, O'Donovan CB, et al. Physical activity attenuates the effect of the FTO genotype on obesity traits in European adults: The Food4Me study. Obesity 2016;24(4):962–9.

109. Livingstone KM, Celis-Morales C, Papandonatos GD, Erar B, Florez JC, Jablonski K a, et al. FTO genotype and weight loss: systematic review and meta-analysis of 9563 individual participant data from eight randomised controlled trials. BMJ 2016;354(4707):1–11.

110. Qi Q, Bray G a, Smith SR, Hu FB, Sacks FM, Qi L. Insulin receptor substrate 1 gene variation modifies insulin resistance response to weight-loss diets in a 2-year randomized trial: the Preventing Overweight Using Novel Dietary Strategies (POUNDS LOST) trial. Circulation 2011;124(5):563–71.

111. Uusitupa M. Gene-diet interaction in relation to the prevention of obesity and type 2 diabetes: evidence from the Finnish Diabetes Prevention Study. Nutr Metab Cardiovasc Dis 2005;15(3):225–33.

112. Florez J, Jablonski K, Bayley N, Pollin T, Bakker PIW De, Ph D, et al. TCF7L2 Polymorphisms and Progression to Diabetes in the Diabetes Prevention Program. N Engl J Med 2006;355(3):241–50.

113. Jablonski KA, Mcateer JB, Bakker PIW De, Franks PW, Pollin TI, Hanson RL, et al. Common Variants in 40 Genes Assessed for Diabetes Incidence and Response to Metformin and Lifestyle Intervention in the Diabetes Prevention Program. Diabetes 2010;59:2672–81.

114. Qi B, Bray GA, Hu FB, Sacks FM, Qi L. Weight-loss diets modify glucose-dependent insulinotropic polypeptide receptor rs2287019 genotype effects on changes in body weight, fasting glucose, and insulin resistance: The preventing overweight using novel dietary strategies trial. Am J Clin Nutr 2012;95(2):506–13.

115. Carvalho-Wells AL, Jackson KG, Lockyer S, Lovegrove JA, Minihane AM. APOE genotype influences triglyceride and C-reactive protein responses to altered dietary fat intake in UK

adults. Am J Clin Nutr 2012;96(6):1447–53.

116. Fallaize R, Celis-Morales C, Macready AL, Marsaux CF, Forster H, O'Donovan C, et al. The effect of the apolipoprotein E genotype on response to personalized dietary advice intervention: findings from the Food4Me randomized controlled trial. Am J Clin Nutr 2016;104(3):827–36.

117. Harrington JM, Phillips CM. Nutrigenetics: bridging two worlds to understand type 2 diabetes. Curr Diab Rep 2014;14(4):477.

118. Kilpeläinen TO. Common Sources of Bias in Gene–Lifestyle Interaction Studies of Cardiometabolic Disease. Curr Nutr Rep 2013;2(4):251–7.

119. Cornelis MC, Hu FB. Gene-environment interactions in the development of type 2 diabetes: recent progress and continuing challenges. Annu Rev Nutr 2012;32(April):245–59.

120. Aschard H, Lutz S, Maus B, Duell EJ, Fingerlin TE, Chatterjee N, et al. Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. Hum Genet 2012;131(10):1591–613.

121. Luan JA, Wong MY, Day NE, Wareham NJ. Sample size determination for studies of gene-environment interaction. Int J Epidemiol 2001;30(5):1035–40.

122. Wong M, Day N, Luan J, Chan K, Wareham N. The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? Int J Epidemiol 2003;32(1):51–7.

123. Wei W-H, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nat Rev Genet 2014;15(11):722–33.

124. Franks PW. Gene × environment interactions in type 2 diabetes. Curr Diab Rep 2011;11(6):552–61.

125. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet 2013;9(3):e1003348.

126. Sul JH, Bilow M, Yang W-Y, Kostem E, Furlotte N, He D, et al. Accounting for Population Structure in Gene-by-Environment Interactions in Genome-Wide Association Studies Using Mixed Models. PLoS Genet 2016;12(3):e1005849.

127. VanderWeele TJ, Ko YA, Mukherjee B. Environmental confounding in gene-environment interaction studies. Am J Epidemiol 2013;178(1):144–52.

128. Keller MC. Gene × environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. Biol Psychiatry 2014;75(1):18–24.

129. Shieh G. Clarifying the role of mean centring in multicollinearity of interaction effects. Br J Math Stat Psychol 2011;64(3):462–77.

130. Kutz G. Nutrigenetic Testing. Tests Purchased from Four Web Sites Mislead Consumers. USA: 2006.

131. Green R, Farahany N. The FDA is overcautious on consumer genomics. Nature 2014;505:286–7.

132. Hogarth S, Javitt G, Melzer D. The current landscape for direct-to-consumer genetic testing: legal, ethical, and policy issues. Annu Rev Genomics Hum Genet 2008;9:161–82.

133. ACCE Model Process for Evaluating Genetic Tests [Internet]. CDC [cited 2015 May 9];Available from: http://www.cdc.gov/genomics/gtesting/ACCE/

134. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd Ed. USA: Lippincott Williams & Wilkins; 2008.

135. VanderWeele TJ. Explanation in Causal Inference: Methods for Mediation and Interaction. USA: Oxford University Press; 2015.

136. VanderWeele TJ, Robins JM. Four types of effect modification: a classification based on directed acyclic graphs. Epidemiology 2007;18(5):561–8.

137. Knol MJ, Van Der Tweel I, Grobbee DE, Numans ME, Geerlings MI. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int J Epidemiol 2007;36(5):1111–8.

138. Knol MJ, VanderWeele TJ, Groenwold RHH, Klungel OH, Rovers MM, Grobbee DE. Estimating measures of interaction on an additive scale for preventive exposures. Eur J Epidemiol 2011;26(6):433–8.

139. Clayton DG. Prediction and interaction in complex disease genetics: Experience in type 1 diabetes. PLoS Genet 2009;5(7):1–6.

140. Markowitz SB, Levin SM, Miller A, Morabia A. Asbestos, Asbestosis, Smoking, and Lung Cancer. New Findings from the North American Insulator Cohort. Am J Respir Crit Care Med 2013;188(1):90–6.

141. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol 2012;41(2):514–20.

142. Cornelis MC. Gene-Diet Interactions in Type 2 Diabetes. Curr Nutr Rep 2014;3(4):302–23.

143. Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, von Elm E, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE Statement. Hum Genet 2009;125(2):131–51.

144. Boffetta P, Winn DM, Ioannidis JP, Thomas DC, Little J, Smith GD, et al. Recommendations and proposed guidelines for assessing the cumulative evidence on joint effects of genes and environments on cancer occurrence in humans. Int J Epidemiol 2012;41(May):686–704.

145. Little J, Manolio T, Melzer D, Minelli C, Pharoah P, Sanderson S, et al. The HuGENet ™ HuGE Review Handbook , version 1.0. 2006.

146. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA 2008;299(11):1335–44.

147. Holmes M V., Newcombe P, Hubacek JA., Sofat R, Ricketts SL, Cooper J, et al. Effect modification by population dietary folate on the association between MTHFR genotype, homocysteine, and stroke risk: A meta-analysis of genetic studies and randomised trials. Lancet 2011;378(9791):584–94.

148. Qi Q, Chu AY, Kang JH, Jensen MK, Curhan GC, Pasquale LR, et al. Sugar-sweetened beverages and genetic risk of obesity. N Engl J Med 2012;367(15):1387–96.

149. Dudbridge F. Polygenic Epidemiology Genetic Epidemiology. Genet Epidemiol 2016;40:268–72.

150. Aschard H. A perspective on interaction effects in genetic association studies. Genet Epidemiol 2016;40(8):678–88.

151. Gong J, Hutter CM, Newcomb PA, Ulrich CM, Bien SA, Campbell PT, et al. Genome-Wide Interaction Analyses between Genetic Variants and Alcohol Consumption and Smoking for Risk of Colorectal Cancer. PLoS Genet 2016;12(10):e1006296.

152. Budin-Ljøsne I, Soye KJ, Tassé AM, Knoppers BM, Harris JR. Genotype-driven recruitment:

a strategy whose time has come? BMC Med Genomics 2013;6(1):19.

153. Nettleton JA, Follis JL, Ngwa JS, Smith CE, Ahmad S, Tanaka T, et al. Gene × dietary pattern interactions in obesity: Analysis of up to 68 317 adults of European ancestry. Hum Mol Genet 2015;24(16):4728–38.

154. Palla L, Higgins JPT, Wareham NJ, Sharp SJ. Challenges in the use of literature-based meta-analysis to examine gene-environment interactions. Am J Epidemiol 2010;171(11):1225–32.

155. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 2002;5(6B):1113–24.

156. Langenberg C, Sharp S, Forouhi NG, Franks PW, Schulze MB, Kerrison N, et al. Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. Diabetologia 2011;54(9):2272–82.

157. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 1986;73(1):1–11.

158. Slimani N, Deharveng G, Unwin I, Southgate DAT, Vignat J, Skeie G, et al. The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. Eur J Clin Nutr 2007;61(9):1037–56.

159. Kroke A, Klipstein-grobusch K, Voss S, Möseneder J, Thielecke F, Noack R, et al. Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water,. Am J Clin Nutr 1999;70:439–47.

160. Bingham S, Welch A, Day K, Cassidy A. Comparison of dietary assessment methods in nutritional epidemiology: weighed Records V. 24 H Recalls , Food-Frequency and estimated-diet records. Br J Nutr 1994;72:619–43.

161. Bingham S, Cassidy A, Cole TJ, Welch A, Runswick SA, Black AE, et al. Validation of weighed records and other methods of dietary assessment using the 24 h urine nitrogen technique and other biological markers. Br J Nutr 1995;73:531–50.

162. Forouhi NG, Koulman A, Sharp SJ, Imamura F, Kröger J, Schulze MB, et al. Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study. Lancet Diabetes Endocrinol 2014;8587(14):1–9.

163. Wang LY, Summerhill K, Rodriguez-Canas C, Mather I, Patel P, Eiden M, et al. Development and validation of a robust automated analysis of plasma phospholipid fatty acids for metabolic phenotyping of large epidemiological studies. Genome Med 2013;5(4):39.

164. Onland-Moret NC, van der A DL, van der Schouw YT, Buschers W, Elias SG, van Gils CH, et al. Analysis of case-cohort data: a comparison of different methods. J Clin Epidemiol 2007;60(4):350–5.

165. Geneletti S, Gallo V, Porta M, Khoury MJ, Vineis P. Assessing causal relationships in genomics: From Bradford-Hill criteria to complex gene-environment interactions and directed acyclic graphs. Emerg Themes Epidemiol 2011;8(1):5.

166. Kirkwood BR, Sterne JA. Essential Medical Statistics. Oxford, UK: Blackwell Publishing

Company; 2013.

167. Rhee JJ, Sampson L, Cho E, Hughes MD, Hu FB, Willett WC. Comparison of methods to account for implausible reporting of energy intake in epidemiologic studies. Am J Epidemiol 2015;181(4):225–33.

168. Willett WC. Nutritional epidemiology. 3rd ed. New York: Oxford University Press; 2013.

169. Zheng J-S, Arnett DK, Lee Y-C, Shen J, Parnell LD, Smith CE, et al. Genome-wide contribution of genotype by environment interaction to variation of diabetes-related traits. PLoS One 2013;8(10):e77442.

170. Wirstrom T, Hilding A, Hu H, Ostenson C-G, Bjorklund A. Consumption of whole grain reduces risk of deteriorating glucose tolerance, including progression to prediabetes. Am J Clin Nutr 2013;97:179–87.

171. Fisher E, Boeing H, Fritsche A, Doering F, Joost H-G, Schulze MB. Whole-grain consumption and transcription factor-7-like 2 ( *TCF7L2*) rs7903146: gene-diet interaction in modulating type 2 diabetes risk. Br J Nutr 2009;101(4):478–81.

172. Cornelis MC, Qi L, Kraft P, Hu FB. *TCF7L2* , dietary carbohydrate and risk of type 2 diabetes in US. Am J Clin Nutr 2009;89:1256–62.

173. Franks PW, Mesa J-L, Harding AH, Wareham NJ. Gene-lifestyle interaction on risk of type 2 diabetes. Nutr Metab Cardiovasc Dis 2007;17(2):104–24.

174. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet 2010;11(7):459–63.

175. Hu FB, Stampfer MJ, Rimm E, Ascherio A, Rosner BA, Spiegelman D, et al. Dietary fat and coronary heart disease: a comparison of approaches for adjusting for total energy intake and modeling repeated dietary measurements. Am J Epidemiol 1999;149(6):531–40.

176. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of Observational Studies in Epidemiology A proposal for reporting. JAMA 2000;283(15):2008–12.

177. Stewart L, Moher D, Shekelle P. Why prospective registration of systematic reviews makes sense. Syst Rev 2012;1(7):1–4.

178. Lunn J, Theobald H. The health effects of dietary unsaturated fatty acids. 2006.

179. SACN. Draft Carbohydrates and Health report. 2014.

180. Fardet A. New hypotheses for the health-protective mechanisms of whole-grain cereals: what is beyond fibre? Nutr Res Rev 2010;23(1):65–134.

181. Moura FF De, Lewis KD, Falk MC. Applying the FDA Definition of Whole Grains to the Evidence for Cardiovascular Disease Health Claims. J Nutr 2009;139:2220S–2226S.

182. Corella D, Tai ES, Sorlí J V, Chew SK, Coltell O, Sotos-prieto M, et al. Association between the APOA2 promoter polymorphism and body-weight in Mediterranean and Asian populations. Replication of a gene-saturated fat interaction. Int J Obes (Lond) 2011;35(5):666–75.

183. Pisabarro RE, Sanguinetti C, Stoll M, Prendez D. High Incidence of Type 2 Diabetes in Peroxisome Proliferator – Activated Receptor y2 Pro12Ala Carriers Exposed to a High Chronic Intake of Trans Fatty Acids and Saturated Fatty Acids. Diabetes Care 2004;27(9):2251–2.

184. Sterne JA, Higgins JPT, Reeves BC. A Cochrane Risk Of Bias Assessment Tool : for Non-Randomized Studies of Interventions ( ACROBAT-NRSI ). 2014.

185. Salanti G, Amountza G, Ntzani EE, Ioannidis JP a. Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. Eur J Hum Genet 2005;13(7):840–8.

186. Fisher E, Schreiber S, Joost H, Boeing H, Do F. A Two-Step Association Study Identifies CAV2 rs2270188 Single Nucleotide Polymorphism Interaction with Fat Intake in Type 2 Diabetes Risk. J Nutr 2011;141:177–81.

187. Slimani N, Bingham S, Runswick S, Ferrari P, Day NE, Welch A., et al. Group level validation of protein intakes estimated by 24-hour diet recall and dietary questionnaires against 24-hour urinary nitrogen in the European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study. Cancer Epidemiol Biomarkers Prev 2003;12(8):784–95.

188. Hoffmann K, Kroke A, Klipstein-Grobusch K, Boeing H. Standardization of dietary intake measurements by nonlinear calibration using short-term reference data. Am J Epidemiol 2002;156(9):862–70.

189. Villegas R, Goodloe RJ, McClellan BE, Boston J, Crawford DC. Gene-carbohydrate and gene-fiber interactions and type 2 diabetes in diverse populations from the National Health and Nutrition Examination Surveys ( NHANES ) as part of the Epidemiologic Architecture for Genes Linked to Environment ( EAGLE ) study. BMC Genet 2014;15(69):1–8.

190. Zheng J-S, Arnett DK, Parnell LD, Smith CE, Li D, Borecki IB, et al. Modulation by dietary fat and carbohydrate of IRS1 association with type 2 diabetes traits in two populations of different ancestries. Diabetes Care 2013;36(9):2621–7.

191. Mansego ML, Martínez F, Martínez-Larrad MT, Zabena C, Rojo G, Morcillo S, et al. Common variants of the liver fatty acid binding protein gene influence the risk of type 2 diabetes and insulin resistance in Spanish population. PLoS One 2012;7(3):e31853.

192. Zheng J-S, Huang T, Li K, Chen Y, Xie H, Xu D, et al. Modulation of the Association between the PEPD Variant and the Risk of Type 2 Diabetes by n-3 Fatty Acids in Chinese Hans. J Nutrigenet Nutrigenomics 2015;8(1):36–43.

193. Nelson TL, Fingerlin TE, Moss LK, Barmada MM, Ferrell RE, Norris JM. The peroxisome proliferator-activated receptor gamma coactivator-1 Alpha Gene (PGC-1a) is not associated with Type 2 diabetes mellitus or body mass index among hispanic and non hispanic whites from Colorado. Exp Clin Endocrinol diabetes 2007;115:268–75.

194. Sonestedt E, Lyssenko V, Ericson U, Gullberg B, Wirfält E, Groop L, et al. Genetic variation in the glucose-dependent insulinotropic polypeptide receptor modifies the association between carbohydrate and fat intake and risk of type 2 diabetes in the Malmo Diet and Cancer cohort. J Clin Endocrinol Metab 2012;97(5):E810-8.

195. Ericson U, Rukh G, Stojkovic I, Sonestedt E, Gullberg B, Wirfa E, et al. Sex-specific interactions between the IRS1 polymorphism and intakes of carbohydrates and fat on incident type 2 diabetes. Am J Clin Nutr 2013;97:208–16.

196. Lamri A, Abi Khalil C, Jaziri R, Velho G, Lantieri O, Vol S, et al. Dietary fat intake and polymorphisms at the PPARG locus modulate BMI and type 2 diabetes risk in the D.E.S.I.R. prospective study. Int J Obes 2012;36(2):218–24.

197. Cho YS, Chen C-H, Hu C, Long J, Ong RTH, Sim X, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet 2012;44(1):67–72.

198. Nelson TL, Fingerlin TE, Moss LK, Barmada MM, Ferrell RE, Norris JM. Association of the peroxisome proliferator-activated receptor gamma gene with type 2 diabetes mellitus

varies by physical activity among non-Hispanic whites from Colorado. Metabolism 2007;56(3):388–93.

199. Simonds NI, Ghazarian AA, Pimentel CB, Schully SD, Ellison GL, Gillanders EM, et al. Review of the Gene-Environment Interaction Literature in Cancer: What Do We Know? Genet Epidemiol 2016;40(5):356–65.

200. Franks PW, Shungin D. The interplay of lifestyle and genetic susceptibility in Type 2 diabetes risk. Diabetes Manag 2011;1(3):299–307.

201. Hoggart CJ, Clark TG, Iorio M De, Whittaker JC, Balding DJ. Genome-Wide Significance for Dense SNP and Resequencing Data. Genet Epidemiol 2008;32:179–85.

202. Hutter CM, Mechanic LE, Chatterjee N, Kraft P, Gillanders EM, Abnet CC, et al. Gene-environment interactions in cancer epidemiology: A national cancer institute think tank report. Genet Epidemiol 2013;37:643–57.

203. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) 2005;95(3):221–7.

204. Knol MJ, Le Cessie S, Algra A, Vandenbroucke JP, Groenwold RHH. Overestimation of risk ratios by odds ratios in trials and cohort studies: Alternatives to logistic regression. CMAJ 2012;184(8):895–9.

205. Cornelis MC, Tchetgen EJT, Liang L, Qi L, Chatterjee N, Hu FB, et al. Gene-Environment Interactions in Genome-Wide Association Studies : A Comparative Study of Tests Applied to Empirical Studies of Type 2 Diabetes. Am J Epidmiol 2012;175(3):191–202.

206. Midthune D, Carroll RJ, Freedman LS, Kipnis V. Measurement error models with interactions. Biostatistics 2016;17:277–90.

207. Scott RA, Chu AC, Grarup N, Manning AK, Hivert M-F, Shungin D, et al. No interactions between previously associated 2-hour glucose gene variants and physical activity or BMI on 2-hour glucose levels. Diabetes 2012;61(5):1291–6.

208. Van Horn L, Carson JAS, Appel LJ, Burke LE, Economos C, Karmally W, et al. Recommended Dietary Pattern to Achieve Adherence to the American Heart Association/American College of Cardiology (AHA/ACC) Guidelines: A Scientific Statement from the American Heart Association. Circulation 2016;134(22):e505–29.

209. Li SX, Imamura F, Ye Z, Schulze MB, Zheng J. Interaction between genes and macronutrient intake on the risk of developing type 2 diabetes: systematic review and findings from EPIC- InterAct. Am J Clin Nutr 2017;106(1):263–75.

210. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011;30(July 2010):377–99.

211. Sterne J a C, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338(July):b2393.

212. Sluijs I, Beulens JWJ, Schouw YT Van Der, A DL Van Der, Buckland G, Kuijsten A, et al. Dietary Glycemic Index , Glycemic Load, and Digestible Carbohydrate Intake Are Not Associated with Risk of Type 2 Diabetes in Eight European Countries. J Nutr 2013;143:93–9.

213. van Nielen M, Feskens EJM, Mensink M, Sluijs I, Molina E, Amiano P, et al. Dietary protein intake and incidence of type 2 diabetes in Europe: the EPIC-InterAct Case-Cohort Study. Diabetes Care 2014;37(7):1854–62.

214. Rukh G, Sonestedt E, Melander O, Hedblad B, Wirfält E, Ericson U, et al. Genetic susceptibility to obesity and diet intakes: Association and interaction analyses in the

Malmö Diet and Cancer Study. Genes Nutr 2013;8(6):535–47.

215. Ankarfeldt MZ, Larsen SC, Ängquist L, Husemoen LLN, Roswall N, Overvad K, et al. Interaction between genetic predisposition to adiposity and dietary protein in relation to subsequent change in body weight and waist circumference. PLoS One 2014;9(10):e110890.

216. Goni L, Cuervo M, Milagro FI, Martínez JA. A genetic risk tool for obesity predisposition assessment and personalized nutrition implementation based on macronutrient intake. Genes Nutr 2015;10(1):445.

217. Wang T, Huang T, Zheng Y, Rood J, Bray GA, Sacks FM, et al. Genetic variation of fasting glucose and changes in glycemia in response to 2-year weight-loss diet intervention: the POUNDS LOST trial. Int J Obes 2016;40(7):1164–9.

218. Dudbridge F, Fletcher O. Gene-environment dependence creates spurious gene-environment interaction. Am J Hum Genet 2014;95(3):301–7.

219. Jacobs DR, Gross MD, Tapsell LC. Food synergy: An operational concept for understanding nutrition. In: Am J Clin Nutr. 2009. page 1543–8.

220. Brunkwall L, Chen Y, Hindy G, Rukh G, Ericson U, Barroso I, et al. Sugar-sweetened beverage consumption and genetic predisposition to obesity in 2 Swedish cohorts. Am J Clin Nutr 2016;104(3):809–15.

221. The InterAct Consortium. Investigation of gene–diet interactions in the incretin system and risk of type 2 diabetes: the EPIC-InterAct study. Diabetologia 2016;59(12):2613–21.

222. VanderWeele TJ, Richardson TS. General theory for interactions in sufficient cause models with dichotomous exposures. Ann Stat 2012;40(4):2128–61.

223. Patel PS, Sharp SJ, Jansen E, Luben RN, Khaw K, Wareham NJ, et al. Fatty acids measured in plasma and erythrocyte-membrane phospholipids and derived by food-frequency questionnaire and the risk of new-onset type 2 diabetes : a pilot study in the European Prospective Investigation into Cancer and Nutrition ( EPIC )–Norfolk. Am J Clin Nutr 2010;92:1214–22.

224. Kröger J, Zieteman V, Enzenback C, Weikert C, Jansen EHJM, Döring F, et al. Erythrocyte membrane phospholipid fatty acids , desaturase activity , and dietary fatty acids in relation to risk of type 2 diabetes in the European Prospective Investigation into Cancer and nutrition (EPIC)-Potsdam Study. Am J Clin Nutr 2011;93(1):127–42.

225. Qi L, Cornelis MC, Zhang C, Dam RM Van, Hu FB. Genetic predisposition , Western dietary pattern , and the risk of type 2 diabetes in men. Am J Clin Nutr 2009;89:1453–8.

226. Bingham S, Day N. Using biochemical markers to assessthe validity of prospective dietary assessment methods and the effect of energy adjustment. Am J Clin Nutr 1997;65:1130S–7S.

227. Food Standards Agency. Food portion sizes. 2nd ed. UK: HMSO; 2002.

228. Newson R, ALSPAC study team. Multiple-test procedures and smile plots. Stata J 2003;3(2):109–32.

229. Schwingshackl L, Schwedhelm C, Hoffmann G, Lampousi A-M, Knüppel S, Iqbal K, et al. Food groups and risk of type 2 diabetes mellitus: a systematic review and meta-analysis of prospective studies. Eur J Epidemiol 2017;32:363–75.

230. Sluijs I, Forouhi NG, Beulens JWJ, Schouw YT Van Der, Agnoli C, Arriola L, et al. The amount and type of dairy product intake and incident type 2 diabetes : results from the EPIC-InterAct Study. Am J Clin Nutr 2012;96:382–90.

231. Bendinelli B, Palli D, Masala G, Sharp SJ, Schulze MB, Guevara M, et al. Association between dietary meat consumption and incident type 2 diabetes: The EPIC-InterAct study. Diabetologia 2013;56(1):47–59.

232. Patel PS, Forouhi NG, Kuijsten A, Schulze MB, Woudenbergh GJ Van, Ardanaz E, et al. The prospective association between total and type of fish intake and type 2 diabetes in 8 European countries : EPIC-InterAct Study. Am J Clin Nutr 2012;95:1445–53.

233. The InterAct Consortium. Consumption of sweet beverages and type 2 diabetes incidence in European adults: results from EPIC-InterAct. Diabetologia 2013;56(7):1520–30.

234. Li M, Fan Y, Zhang X, Hou W, Tang Z. Fruit and vegetable intake and risk of type 2 diabetes mellitus: meta-analysis of prospective cohort studies. BMJ Open 2014;4(11):e005497.

235. Qi Q, Chu AY, Kang JH, Liang L, Curhan GC, Pasquale LR, et al. Fried food consumption , genetic risk , and body mass index : gene-diet interaction analysis in three US cohort. 2014;1610(March):1–12.

236. Khera A V., Emdin C a., Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med 2016;(November):NEJMoa1605086.

237. Olsen NJ, A ngquist L, Larsen SC, Linneberg A, Skaaby T, Husemoen LLN, et al. Interactions between genetic variants associated with adiposity traits and soft drinks in relation to longitudinal changes in body weight and waist circumference. Am J Clin Nutr 2016;104(3):816–26.

238. Halliday JL. Genetics and public health--evolution, or revolution? J Epidemiol Community Heal 2004;58(11):894–9.

239. Langenberg C, Sharp SJ, Franks PW, Scott R a, Deloukas P, Forouhi NG, et al. Gene-Lifestyle Interaction and Type 2 Diabetes: The EPIC InterAct Case-Cohort Study. PLoS Med 2014;11(5):e1001647.

240. Pasquale LR, Loomis SJ, Aschard H, Kang JH, Cornelis MC, Qi L, et al. Exploring genome-wide - dietary heme iron intake interactions and the risk of type 2 diabetes. Front Genet 2013;4:7.

241. Figueiredo JC, Hsu L, Hutter CM, Lin Y, Campbell PT, Baron J a., et al. Genome-Wide Diet-Gene Interaction Analyses for Risk of Colorectal Cancer. PLoS Genet 2014;10(4).

242. Melkonian SC, Daniel CR, Hildebrandt M a T, Tannir NM, Ye Y, Chow WH, et al. Joint association of genome-wide association study-identified susceptibility loci and dietary patterns in risk of renal cell carcinoma among non-hispanic whites. Am J Epidemiol 2014;180(5):499–507.

243. Manning AK, Hivert M, Scott R a., Grimsby JL, Bouatia-Naji N, Chen H. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 2012;44(6):659–69.

244. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. Hum Hered 2007;63(2):111–9.

245. Sung YJ, Winkler TW, Manning AK, Aschard H, Gudnason V, Harris TB, et al. An Empirical Comparison of Joint and Stratified Frameworks for Studying G × E Interactions: Systolic Blood Pressure and Smoking in the CHARGE Gene-Lifestyle Interactions Working Group. Genet Epidemiol 2016;40:404–15.

246. Hsu L, Jiao S, Dai JY, Hutter C, Peters U, Kooperberg C. Powerful Cocktail Methods for Detecting Genome-wide Gene- Environment Interaction. Genet Epidemiol

2012;36(3):183–94.

247. Yang J, Loos RJF, Powell JE, Medland SE, Speliotes EK, Chasman DI, et al. FTO genotype is associated with phenotypic variability of body mass index. Nature 2012;490(7419):267–72.

248. Alhazmi A, Stojanovski E, McEvoy M, Garg ML. Macronutrient intake and type 2 diabetes risk in middle-aged Australian women. Results from the Australian Longitudinal Study on Women's Health. Public Health Nutr 2014;17(7):1587–94.

249. Wolfram G, Bechthold A, Boeing H, Ellinger S, Hauner H, Kroke A, et al. Evidence-Based Guideline of the German Nutrition Society: Fat Intake and Prevention of Selected Nutrition-Related Diseases. Ann Nutr Metab 2015;67(3):141–204.

250. Hooper L, Martin N, Abdelhamid A, G DS. Reduction in saturated fat intake for cardiovascular disease (Review). Cochrane Database Syst Rev 2015;(6):1–150.

251. Beulens JWJ, van der Schouw YT, Bergmann MM, Rohrmann S, Schulze MB, Buijsse B, et al. Alcohol consumption and risk of type 2 diabetes in European men and women: influence of beverage type and body size The EPIC-InterAct study. J Intern Med 2012;272(4):358–70.

252. Voorman A, Lumley T, McKnight B, Rice K. Behavior of QQ-plots and Genomic Control in studies of gene-environment interaction. PLoS One 2011;6(5):e19416.

253. EPSRC. Six High Performance Computing centres to be officially launched [Internet]. [cited 2017 Aug 2];Available from: https://www.epsrc.ac.uk/newsevents/news/sixhpccentresofficiallylaunch/

254. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26(17):2190–1.

255. Dadd T, Weale ME, Lewis CM. A critical evaluation of genomic control methods for genetic association studies. Genet Epidemiol 2009;33(4):290–8.

256. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc 2014;9(5):1192–212.

257. Waldman ID. Gene-environment interactions reexamined: Does mother's marital stability interact with the dopamine receptor D2 gene in the etiology of childhood attention-deficit/hyperactivity disorder? Dev Psychopathol 2007;19(2007):1117–28.

258. Day FR, Loh PR, Scott R a., Ong KK, Perry JRB. A Robust Example of Collider Bias in a Genetic Association Study. Am J Hum Genet 2016;98(2):392–3.

259. Fan Q, Verhoeven VJM, Wojciechowski R, Barathi V a, Hysi PG, Guggenheim J a, et al. Meta-analysis of gene-environment-wide association scans accounting for education level identifies additional loci for refractive error. Nat Commun 2016;7:11008.

260. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nat Rev Genet 2010;11(12):843–54.

261. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45(6):580–5.

262. Acosta PB, editor. Nutritional management of patients with inherited metabolic disorders. USA: Jones and Bartlett Publishers; 2010.

263. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012;40(Database issue):D109-14.

264. Hochberg Y, Benjaminit Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Source J R Stat Soc 1995;57(1):289–300.

265. Delgado-Rodriguez M. Bias. J Epidemiol Community Heal 2004;58(8):635–41.

266. Teixeira da Silva JA. Negative results: negative perceptions limit their potential for increasing reproducibility. J Negat Results Biomed 2015;14(12):1–4.

267. Tabery J. Beyond Versus: The struggle to Understand the Interaction of Nature and Nurture. USA: MIT Press; 2014.

268. Tabery J. R. A. Fisher, Lancelot Hogben, and the Origin(s) of Genotype-Environment Interaction. J Hist Biol 2008;41:717–61.

269. Tabery J. Debating interaction: the history, and an explanation. Int J Epidemiol 2015;44(4):1117–23.

270. Ortlipp M. Keeping and Using Reflective Journals in the Qualitative Research Process. Qual Rep 2008;13(4):695–705.

271. Vimaleswaran KS, Bodhini D, Lakshmipriya N, Ramya K, Anjana RM, Sudha V, et al. Interaction between FTO gene variants and lifestyle factors on metabolic traits in an Asian Indian population. Nutr Metab (Lond) 2016;13:39.

272. Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965;58:295–300.

273. Lucas RM, McMichael AJ. Association or causation: evaluating links between “environment and disease”. Bull World Health Organ 2005;83(10):792–5.

274. Balansky R, Ganchev G, Iltcheva M, Nikolov M, Steele VE, De Flora S. Differential carcinogenicity of cigarette smoke in mice exposed either transplacentally, early in life or in adulthood. Int J Cancer 2012;130(5):1001–10.

275. Lo C-Y, Hsieh P-H, Chen H-F, Su H-M. A maternal high-fat diet during pregnancy in rats results in a greater risk of carcinogen-induced mammary tumors in the female offspring than exposure to a high-fat diet in postnatal life. Int J cancer 2009;125(4):767–73.

276. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HYK, Chen R, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 2012;148(6):1293–307.

277. Wenten M, Gauderman WJ, Berhane K, Lin PC, Peters J, Gilliland FD. Functional variants in the catalase and myeloperoxidase genes, ambient air pollution, and respiratory-related school absences: An example of epistasis in gene-environment interactions. Am J Epidemiol 2009;170(12):1494–501.

278. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 2001;69(1):138–47.

279. Manuck SB, McCaffery JM. Gene-Environment Interaction. Annu Rev Psychol 2014;65(1):41–70.

280. Strecher VJ, Rosenstock IM. The health belief model. In: Cambridge handbook of psychology, health and medicine. Cambridge: Cambridge University Press; 1997. page 113–117.

281. Schneider KI, Schmidtke J. Patient compliance based on genetic medicine: a literature review. J Community Genet 2014;5(1):31–48.

282. Li SX, Ye Z, Whelan K, Truby H. The effect of communicating the genetic risk of cardiometabolic disorders on motivation and actual engagement in preventative lifestyle

modification and clinical outcome: a systematic review and meta-analysis of randomised controlled trials. Br J Nutr 2016;116(5):924–34.

283. Hollands GJ, French DP, Griffin SJ, Prevost AT, Sutton S, King S, et al. The impact of communicating genetic risks of disease on risk- reducing health behaviour : systematic review with meta-analysis. BMJ 2016;352(1102):1–11.

284. Celis-Morales C, Livingstone KM, Marsaux CFM, Macready AL, Fallaize R, O'Donovan CB, et al. Effect of personalized nutrition on health-related behaviour change: evidence from the Food4me European randomized controlled trial. Int J Epidemiol 2017;46(2):578–88.

285. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Smith GD. Best (but oft-forgotten) practices: The design, analysis, and interpretation of Mendelian randomization studies. Am J Clin Nutr 2016;103(4):965–78.

286. Burgess S, Butterworth A, Malarstig A, Thompson SG. Use of Mendelian randomisation to assess potential benefit of clinical intervention. BMJ 2012;345(November):1–6.

287. Qi Q, Kilpeläinen TO, Downer MK, Tanaka T, Smith CE, Sluijs I, et al. FTO genetic variants, dietary intake, and body mass index: insights from 177,330 individuals. Hum Mol Genet 2014;1–12.

288. Fortier I, Burton PR, Robson PJ, Ferretti V, Little J, L'Heureux F, et al. Quality, quantity and harmony: The DataSHaPER approach to integrating data across bioclinical studies. Int J Epidemiol 2010;39(5):1383–93.

289. Fortier I, Raina P, Van den Heuvel ER, Griffith LE, Craig C, Saliba M, et al. Maelstrom Research guidelines for rigorous retrospective data harmonization. Int J Epidemiol 2017;46(1):103–15.

290. Smith GD. Mendelian Randomization for Strengthening Causal Inference in Observational Studies: Application to Gene x Environment Interactions. Perspect Psychol Sci 2011;6(3):314–314.

# *Appendices*

## *Appendix A: Checklist for critiquing gene-diet interaction studies*

Purpose: to highlight the key concepts within a research paper about gene-diet interactions or nutrigenetics that readers need to be mindful of when assessing the quality of the study.

Additionally, like any epidemiological study, the validity of the study depends on whether it is robust to confounding, bias and/or chance (see STROBE and/or Cochrane RCT quality checklists).

| **Concept** | **Rationale** | **Assessment notes** |
|---|---|---|
| Study design: if it is prospective or retrospective | Retrospective study designs e.g. cross-sectional or case-control studies make it difficult to determine the temporal sequence of events and therefore whether diet led to disease or the other way around (reverse causality). More about the general issues relating to different study designs can be found in textbooks on epidemiology. | |
| Sample size large enough? | A key determinant of a GWAS.<br><br>Common genetic variants (that occur frequently in the population) often have small effects on the outcome of interest for polygenic conditions and therefore studies with large sample sizes are needed.<br><br>If available, a sample size calculation may help you determine whether the sample size was large enough. | |
| How well is the dietary exposure measured? | Is it a subjective (e.g. self-report) or objective (e.g. nutritional biomarker, via observation or dietary intervention) measure of dietary intake?<br><br>Has this tool been validated? | |
| How good is the genotyping and imputation, where relevant? (quality control of the study) | Look for information about genotyping performance. The sample and SNP call rate should be >95%. Those <95% are normally removed from further analysis because of the possibility of high genotyping errors contributing to bias.<br><br>More on this can be found in the paper by Pearson, JAMA, 2008<br><br>If some SNPs are imputed (estimated from the genotyped SNPs), authors should provide the average quality score or information metric. | |
| Appropriateness of interaction scale: additive or multiplicative interaction? | An interaction on the multiplicative scale does not imply one on the additive scale. Vice versa. Also, a lack of interaction on the multiplicative scale does not preclude a lack of interaction on the additive scale.<br><br>Most interaction studies are under the multiplicative scale because of ease of analysis.<br><br>Additive interactions should be investigated to understand an issue of public health relevance. Whereas a multiplicative interaction | |

| | | |
|---|---|---|
| | should be assessed where such relevance may not yet be so obvious but may serve to help understand disease aetiology. | |
| Have important confounders been addressed? | For interaction analyses, it is important to control for two sets of confounders related to genes and disease as well as to diet and disease. Therefore, for the latter, confounders may include physical activity, socio-economic status (e.g. education level), smoking, alcohol intake and total energy intake<br><br>An important confounder for gene-disease associations is population stratification (or population structure: how ethnically diverse is the population under study)? This is often adjusted in the analysis using eigenvectors or principal components for population stratification.<br><br>Often authors will report whether their genotypes are in Hardy-Weinberg equilibrium. If not (with p-value statistically significantly deviating from equilibrium), this may indicate genotyping or genotype calling errors or peculiarities in the data. | |
| Were stringent corrections applied to multiple testing? | To reduce the likelihood of false positive/chance findings, often a problem with multiple tests, Bonferroni correction is commonly done. However, other less conservative approaches may have also been considered. Authors should have described this in their methods and discussed their results taking this into account. | |
| If there is an interaction, is both the overall p-value for interaction across strata and the strata-specific effect estimates presented? | An interaction is present when there are a statistically significant difference in effect estimates between strata of either genotype or the dietary exposure category (e.g. is there a significant difference between the association between the highest compared to lowest quartile of fibre intake and disease?) | |
| Have the results been replicated in an independent population? Is there consistency in findings across different types of studies? | To provide insurance against errors or biases that can affect an individual study, replication of GWAS studies is now mandatory. However, replication for gene-diet interactions is more challenging because of differences in population characteristics and methods in dietary intake. Therefore, whilst it will be preferable that studies also include independent replication, the inclusion of other tests e.g. functional tests can also help elucidate the validity of the interaction. This concurs with the Bradford Hill criteria of 'consistency,' of findings. | |
| Is there evidence of a functional role of the genetic variants? | Many GWAS associated genetic variants have been in non-coding regions of the DNA (introns- regions that do not code for a protein). To understand what its effect on disease is, further studies examining the role of the genetic variant is often conducted.<br><br>Is the interaction biologically plausible? Although given much of human biology is unknown, this needn't be major criteria. | |
| Are the strata-specific effects of clinical value? | Additive interaction may be undertaken to assess this | |
| **Overall** | **Make an overall impression statement, including where the study does well or can see improvement in.** | |

## *Appendix B: Example of a Food Frequency Questionnaire (centre: EPIC-Norfolk)*

# FOOD FREQUENCY QUESTIONNAIRE

**This questionnaire asks for some background information about you, especially about what you eat.**

**Please answer every question. If you are uncertain about how to answer a question then do the best you can, but please do not leave a question blank.**

1. **YOUR DIET LAST YEAR**

For each food there is an amount shown, either a "medium serving" or a common household unit such as a slice or teaspoon. Please put a tick (✓) in the box to indicate how often, **on average**, you have eaten the specified amount of each food **during the past year**.

**EXAMPLES:**

For white bread the amount is one slice, so if you ate 4 or 5 slices a day, you should put a tick in the column headed "4-5 per day".

| FOODS AND AMOUNTS | AVERAGE USE LAST YEAR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **BREAD AND SAVOURY BISCUITS** (one slice or biscuit) | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |
| White bread and rolls | | | | | | | | ✓ | |

For chips, the amount is a "medium serving", so if you had a helping of chips twice a week you should put a tick in the column headed "2-4 per week".

| FOODS AND AMOUNTS | AVERAGE USE LAST YEAR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **POTATOES, RICE AND PASTA** (medium serving) | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |
| Chips | | | | ✓ | | | | | |

For very seasonal fruits such as strawberries and raspberries you should estimate your average use when the fruits are in season, so if you ate strawberries or raspberries about once a week when they were in season you should put a tick in the column headed "once a week"

| FOODS AND AMOUNTS | AVERAGE USE LAST YEAR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **FRUIT** (1 fruit or medium serving) | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |
| Strawberries, raspberries, kiwi fruit | | | ✓ | | | | | | |

**Please estimate your average food use as best you can, and please answer every question - do not leave ANY lines blank. PLEASE PUT A TICK (✓) ON EVERY LINE**

| FOODS AND AMOUNTS | AVERAGE USE LAST YEAR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **MEAT AND FISH** (medium serving) | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |
| Beef: roast, steak, mince, stew or casserole | | | | | | | | | |
| Beefburgers | | | | | | | | | |
| Pork: roast, chops, stew or slices | | | | | | | | | |
| Lamb: roast, chops or stew | | | | | | | | | |
| Chicken or other poultry eg. turkey | | | | | | | | | |
| Bacon | | | | | | | | | |
| Ham | | | | | | | | | |
| Corned beef, Spam, luncheon meats | | | | | | | | | |
| Sausages | | | | | | | | | |
| Savoury pies, eg. meat pie, pork pie, pasties, steak & kidney pie, sausage rolls | | | | | | | | | |
| Liver, liver paté, liver sausage | | | | | | | | | |
| Fried fish in batter, as in fish and chips | | | | | | | | | |
| Fish fingers, fish cakes | | | | | | | | | |
| Other white fish, fresh or frozen, eg. cod, haddock, plaice, sole, halibut | | | | | | | | | |
| Oily fish, fresh or canned, eg. mackerel, kippers, tuna, salmon, sardines, herring | | | | | | | | | |
| Shellfish, eg. crab, prawns, mussels | | | | | | | | | |
| Fish roe, taramasalata | | | | | | | | | |
| | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |

**Please check that you have a tick (✓) on EVERY line**

**PLEASE PUT A TICK (✓) ON EVERY LINE**

| FOODS AND AMOUNTS | AVERAGE USE LAST YEAR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **BREAD AND SAVOURY BISCUITS** (one slice or biscuit) | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |
| White bread and rolls | | | | | | | | | |
| Brown bread and rolls | | | | | | | | | |
| Wholemeal bread and rolls | | | | | | | | | |
| Cream crackers, cheese biscuits | | | | | | | | | |
| Crispbread, eg. Ryvita | | | | | | | | | |
| **CEREALS** (one bowl) | | | | | | | | | |
| Porridge, Readybrek | | | | | | | | | |
| Breakfast cereal such as cornflakes, muesli etc. | | | | | | | | | |
| **POTATOES, RICE AND PASTA** (medium serving) | | | | | | | | | |
| Boiled, mashed, instant or jacket potatoes | | | | | | | | | |
| Chips | | | | | | | | | |
| Roast potatoes | | | | | | | | | |
| Potato salad | | | | | | | | | |
| White rice | | | | | | | | | |
| Brown rice | | | | | | | | | |
| White or green pasta, eg. spaghetti, macaroni, noodles | | | | | | | | | |
| Wholemeal pasta | | | | | | | | | |
| Lasagne, moussaka | | | | | | | | | |
| Pizza | | | | | | | | | |
| | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |

**Please check that you have a tick (✓) on EVERY line**

**PLEASE PUT A TICK (✓) ON EVERY LINE**

| FOODS AND AMOUNTS | AVERAGE USE LAST YEAR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **DAIRY PRODUCTS AND FATS** | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |
| Single or sour cream (tablespoon) | | | | | | | | | |
| Double or clotted cream (tablespoon) | | | | | | | | | |
| Low fat yogurt, fromage frais (125g carton) | | | | | | | | | |
| Full fat or Greek yogurt (125g carton) | | | | | | | | | |
| Dairy desserts (125g carton) | | | | | | | | | |
| Cheese, eg. Cheddar, Brie, Edam (medium serving) | | | | | | | | | |
| Cottage cheese, low fat soft cheese (medium serving) | | | | | | | | | |
| Eggs as boiled, fried, scrambled, etc. (one) | | | | | | | | | |
| Quiche (medium serving) | | | | | | | | | |
| Low calorie, low fat salad cream(tablespoon) | | | | | | | | | |
| Salad cream, mayonnaise (tablespoon) | | | | | | | | | |
| French dressing (tablespoon) | | | | | | | | | |
| Other salad dressing (tablespoon) | | | | | | | | | |
| **The following on bread or vegetables** | | | | | | | | | |
| Butter (teaspoon) | | | | | | | | | |
| Block margarine, eg. Stork, Krona (teaspoon) | | | | | | | | | |
| Polyunsaturated margarine (tub), eg. Flora, sunflower (teaspoon) | | | | | | | | | |
| Other soft margarine,dairy spreads (tub), eg. Blue Band, Clover (teaspoon) | | | | | | | | | |
| Low fat spread (tub), eg. Outline, Gold (teaspoon) | | | | | | | | | |
| Very low fat spread (tub) (teaspoon) | | | | | | | | | |
| | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |

**Please check that you have a tick (✓) on EVERY line**

**PLEASE PUT A TICK (✓) ON EVERY LINE**

| FOODS AND AMOUNTS | AVERAGE USE LAST YEAR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **SWEETS AND SNACKS** (medium serving) | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |
| Sweet biscuits, chocolate , eg. digestive (one) | | | | | | | | | |
| Sweet biscuits, plain, eg. Nice, ginger (one) | | | | | | | | | |
| Cakes eg. fruit, sponge, home baked | | | | | | | | | |
| Cakes eg. fruit, sponge, ready made | | | | | | | | | |
| Buns, pastries eg. scones, flapjacks, home baked | | | | | | | | | |
| Buns, pastries eg. croissants, doughnuts, ready made | | | | | | | | | |
| Fruit pies, tarts, crumbles, home baked | | | | | | | | | |
| Fruit pies, tarts, crumbles, ready made | | | | | | | | | |
| Sponge puddings, home baked | | | | | | | | | |
| Sponge puddings, ready made | | | | | | | | | |
| Milk puddings, eg. rice, custard, trifle | | | | | | | | | |
| Ice cream, choc ices | | | | | | | | | |
| Chocolates, single or squares | | | | | | | | | |
| Chocolate snack bars eg. Mars, Crunchie | | | | | | | | | |
| Sweets, toffees, mints | | | | | | | | | |
| Sugar added to tea, coffee, cereal (teaspoon) | | | | | | | | | |
| Crisps or other packet snacks, eg. Wotsits | | | | | | | | | |
| Peanuts or other nuts | | | | | | | | | |
| **SOUPS, SAUCES, AND SPREADS** | | | | | | | | | |
| Vegetable soups (bowl) | | | | | | | | | |
| Meat soups (bowl) | | | | | | | | | |
| Sauces, eg. white sauce, cheese sauce, gravy (tablespoon) | | | | | | | | | |
| Tomato ketchup (tablespoon) | | | | | | | | | |
| Pickles, chutney (tablespoon) | | | | | | | | | |
| Marmite, Bovril (teaspoon) | | | | | | | | | |
| Jam, marmalade, honey (teaspoon) | | | | | | | | | |
| Peanut butter (teaspoon) | | | | | | | | | |
| | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |

**Please check that you have a tick (✓) on EVERY line**

**PLEASE PUT A TICK (✓) ON EVERY LINE**

| FOODS AND AMOUNTS | AVERAGE USE LAST YEAR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **DRINKS** | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |
| Tea (cup) | | | | | | | | | |
| Coffee, instant or ground (cup) | | | | | | | | | |
| Coffee, decaffeinated (cup) | | | | | | | | | |
| Coffee whitener, eg. Coffee-mate (teaspoon) | | | | | | | | | |
| Cocoa, hot chocolate (cup) | | | | | | | | | |
| Horlicks, Ovaltine (cup) | | | | | | | | | |
| Wine (glass) | | | | | | | | | |
| Beer, lager or cider (half pint) | | | | | | | | | |
| Port, sherry, vermouth, liqueurs (glass) | | | | | | | | | |
| Spirits, eg. gin, brandy, whisky, vodka (single) | | | | | | | | | |
| Low calorie or diet fizzy soft drinks (glass) | | | | | | | | | |
| Fizzy soft drinks, eg. Coca cola, lemonade (glass) | | | | | | | | | |
| Pure fruit juice (100%) eg. orange, apple juice (glass) | | | | | | | | | |
| Fruit squash or cordial (glass) | | | | | | | | | |
| **FRUIT**<br>**For seasonal fruits marked *, please estimate your average use when the fruit is in season** | | | | | | | | | |
| Apples (1 fruit) | | | | | | | | | |
| Pears (1 fruit) | | | | | | | | | |
| Oranges, satsumas, mandarins (1 fruit) | | | | | | | | | |
| Grapefruit (half) | | | | | | | | | |
| Bananas (1 fruit) | | | | | | | | | |
| Grapes (medium serving) | | | | | | | | | |
| Melon (1 slice) | | | | | | | | | |
| * Peaches, plums, apricots (1 fruit) | | | | | | | | | |
| * Strawberries, raspberries, kiwi fruit (medium serving) | | | | | | | | | |
| Tinned fruit (medium serving) | | | | | | | | | |
| Dried fruit, eg. raisins, prunes (medium serving) | | | | | | | | | |
| | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |

**Please check that you have a tick (✓) on EVERY line**

**PLEASE PUT A TICK (✓) ON EVERY LINE**

| FOODS AND AMOUNTS | AVERAGE USE LAST YEAR | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **VEGETABLES** Fresh, frozen or tinned (medium serving) | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |
| Carrots | | | | | | | | | |
| Spinach | | | | | | | | | |
| Broccoli, spring greens, kale | | | | | | | | | |
| Brussels sprouts | | | | | | | | | |
| Cabbage | | | | | | | | | |
| Peas | | | | | | | | | |
| Green beans, broad beans, runner beans | | | | | | | | | |
| Marrow, courgettes | | | | | | | | | |
| Cauliflower | | | | | | | | | |
| Parsnips, turnips, swedes | | | | | | | | | |
| Leeks | | | | | | | | | |
| Onions | | | | | | | | | |
| Garlic | | | | | | | | | |
| Mushrooms | | | | | | | | | |
| Sweet peppers | | | | | | | | | |
| Beansprouts | | | | | | | | | |
| Green salad, lettuce, cucumber, celery | | | | | | | | | |
| Watercress | | | | | | | | | |
| Tomatoes | | | | | | | | | |
| Sweetcorn | | | | | | | | | |
| Beetroot | | | | | | | | | |
| Coleslaw | | | | | | | | | |
| Avocado | | | | | | | | | |
| Baked beans | | | | | | | | | |
| Dried lentils, beans, peas | | | | | | | | | |
| Tofu , soya meat, TVP, Vegeburger | | | | | | | | | |
| | Never or less than once/month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |

**Please check that you have a tick (✓) on EVERY line**

## YOUR DIET LAST YEAR, continued

2. Are there any **OTHER** foods which you ate more than once a week? Yes ☐ No ☐

**If yes,** please list below

| Food | Usual serving size | Number of times eaten each week |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

3. What type of milk did you most often use?

**Select one only**

Full cream, silver ☐
Skimmed/blue ☐
Dried milk ☐
Other, *specify* ☐
Semi-skimmed, red/white ☐
Channel Islands, gold ☐
Soya ☐
None ☐

4. How much milk did you drink each day, including milk with tea, coffee, cereals etc?

None ☐
Quarter of a pint ☐
Half a pint ☐
Three quarters of a pint ☐
One pint ☐
More than one pint ☐

5. Did you usually eat breakfast cereal (excluding porridge and Ready Brek mentioned earlier)? Yes ☐ No ☐

**If yes,** which brand and type of breakfast cereal, including muesli, did you usually eat?

**List the one or two types most often used**

| Brand *e.g. Kellogg's* | Type *e.g. cornflakes* |
|---|---|
| | |
| | |

6. What kind of fat did you most often use for frying, roasting, grilling etc?

**Select one only**

Butter ☐
Lard/dripping ☐
Vegetable oil ☐
Solid vegetable fat ☐
Margarine ☐
None ☐

**If you used vegetable oil**, please give type eg. corn, sunflower ☐

7. What kind of fat did you most often use for baking cakes etc?

**Select one only**

Butter ☐
Lard/dripping ☐
Vegetable oil ☐
Solid vegetable fat ☐
Margarine ☐
None ☐

**If you used margarine**, please give name or type eg. Flora, Stork ☐

8. How often did you eat food that was fried at home?

Daily ☐ 1-3 times a week ☐ 4-6 times a week ☐
Less than once a week ☐ Never ☐

9. How often did you eat fried food away from home?

Daily ☐ 1-3 times a week ☐ 4-6 times a week ☐
Less than once a week ☐ Never ☐

10. What did you do with the visible fat on your meat?

Ate most of the fat ☐ Ate as little as possible ☐
Ate some of the fat ☐ Did not eat meat ☐

11. How often did you eat grilled or roast meat? ☐☐ times a week

12. How well cooked did you usually have grilled or roast meat?

Well done /dark brown ☐ Lightly cooked/rare ☐
Medium ☐ Did not eat meat ☐

13. How often did you add salt to food while cooking?

Always ☐ Rarely ☐
Usually ☐ Never ☐
Sometimes ☐

14. How often did you add salt to any food at the table?

Always ☐ Rarely ☐
Usually ☐ Never ☐
Sometimes ☐

15. Did you regularly use a salt substitute (eg LoSalt)? Yes ☐ No ☐

**If yes**, which brand? ☐

16. During the course of last year, on average, how many times a week did you eat the following foods?

| **Food type** | **Times/week** | **Portion size** |
|---|---|---|
| Vegetables (not including potatoes) | ☐☐ | medium serving |
| Salads | ☐☐ | medium serving |
| Fruit and fruit products (not including fruit juice) | ☐☐ | medium serving or 1 fruit |
| Fish and fish products | ☐☐ | medium serving |
| Meat, meat products and meat dishes (including bacon, ham and chicken) | ☐☐ | medium serving |

17. Have you taken any vitamins, minerals, fish oils, fibre or other food supplements during the past year? Yes ☐ No ☐ Don't know ☐

**If yes,** please complete the table below. If you have taken more than 5 types of supplement please put the most frequently consumed brands first.

| **Vitamin supplements** | **Average frequency** Tick **one** box per line to show how often on average you consumed supplements | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Name and brand** Please list full name, **brand** and strength | **Dose** Please state number of pills, capsules or teaspoons consumed | Never or less than once a month | 1-3 per month | Once a week | 2-4 per week | 5-6 per week | Once a day | 2-3 per day | 4-5 per day | 6+ per day |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |

*Thank you for your help*

## *Appendix C: Example of a search strategy for the systematic review of gene-macronutrient interactions and T2D*

**EMBASE (Ovid)**
**Date:** 6-10-15

| No. | Query | Result |
|---|---|---|
| **1** | exp Genotype/ | 324549 |
| **2** | exp Alleles/ | 146253 |
| **3** | exp genetic variability/ | 185093 |
| **3** | exp genetic polymorphism/ or exp genetic heterogeneity/ | 361288 |
| **4** | Genetic Predisposition.mp. or exp genetic predisposition/ | 129174 |
| **5** | exp genetic association/ | 119908 |
| **6** | (genotype or 'gene' or 'genetic risk score' or GWAS).ab,ti. | 1948179 |
| **7** | exp genetic polymorphism/ or exp genetic heterogeneity/ | 324073 |
| **8** | 1 or 2 or 3 or 4 or 5 or 6 or 7 | 2368293 |
| **9** | exp non insulin dependent diabetes mellitus/ | 167582 |
| **10** | diabetogenic.mp. | 8417 |
| **11** | 9 or 10 | 175577 |
| **12** | exp carbohydrate diet/ | 14648 |
| **13** | exp glycemic index/ | 3492 |
| **14** | ('glycaemic index' or 'glycemic load' or 'glycaemic load' or GI or GL).ab,ti. | 71757 |
| **15** | exp dietary fiber/ | 16438 |
| **16** | exp sugar intake/ | 4969 |
| **17** | ('free sugar*' or 'added sugar*').ab,ti. | 2039 |
| **18** | exp fat intake/ | 44614 |
| **19** | ('fatty acid*' or 'saturated fat*' or 'monounsaturated fat*' or 'polyunsaturated fat*' or 'trans fat*' or 'dietary cholesterol').ab,ti. | 231922 |
| **20** | ('alpha-linolenic acid' or ALA or 'linolenic acid*' or 'linoleic acid*' or LA or n-3 or n-6 or omega-3 or omega-6).ab,ti. | 321990 |
| **21** | exp protein intake/ | 126 |
| **22** | dietary protein quality.mp. | 100 |
| **23** | exp caloric intake/ | 44733 |
| **24** | exp low carbohydrate diet/ or exp carbohydrate intake/ | 19153 |
| **25** | exp atherogenic diet/ or exp high fiber diet/ or exp lipid diet/ or exp low calory diet/ or exp cholesterol diet/ or exp low fat diet/ or exp protein diet/ | 54183 |
| **26** | (macronutrient or 'dietary fat' or 'dietary protein' or 'dietary carbohydrate').ab,ti. | 26536 |
| **27** | 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 | 740148 |
| **28** | 8 and 11 and 27 | 2152 |
| **29** | **limit 28 to human** | **1509** |

## *Appendix D: EPIC-InterAct gene-macronutrient interactions and incident T2D, under the 'Modified Analyses'*

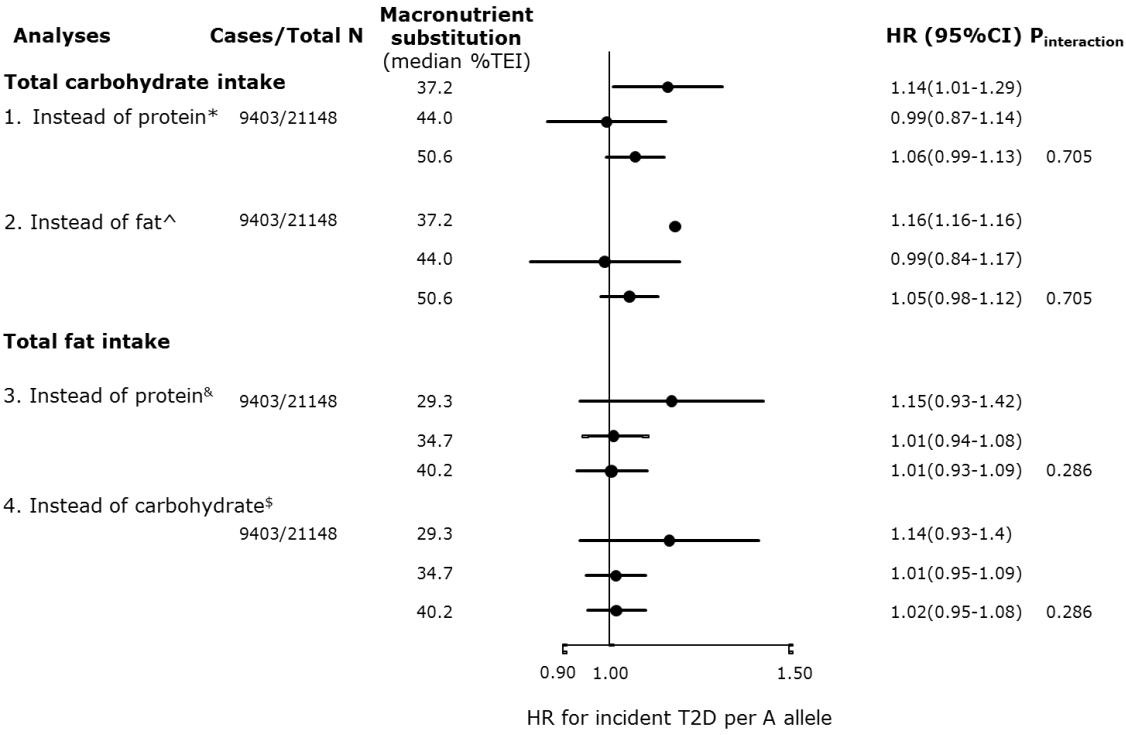



**Figure 1: EPIC-InterAct analyses for the hazard ratio (HR) for incident T2D per A allele (*GIPR* rs10423928), by categories of macronutrient exchange**

Pooled HR for T2D, with adjustment for age (=underlying time scale), sex, centre, eigenvectors (first 5 principal components for population stratification), physical activity, education, smoking, sex-specific alcohol categories, total energy intake (TEI), season, dietary factors (fibre, magnesium, iron, vitamin C, coffee, tea, leafy vegetables and artificially sweetened beverage) and BMI.

Models substituted for the following were further adjusted for other macronutrients* SFA, MUFA and PUFA; ^ protein; & carbohydrate; $ protein

$P_{interaction}$: estimated by treating rs10423928 and macronutrients as continuous variables

Heterogeneity $I^2$ between countries was low to moderate: total Carbohydrate: 0% and total fat: 29%.

Statistical test: multiplicative interaction analysis using Prentice-weighted Cox regression

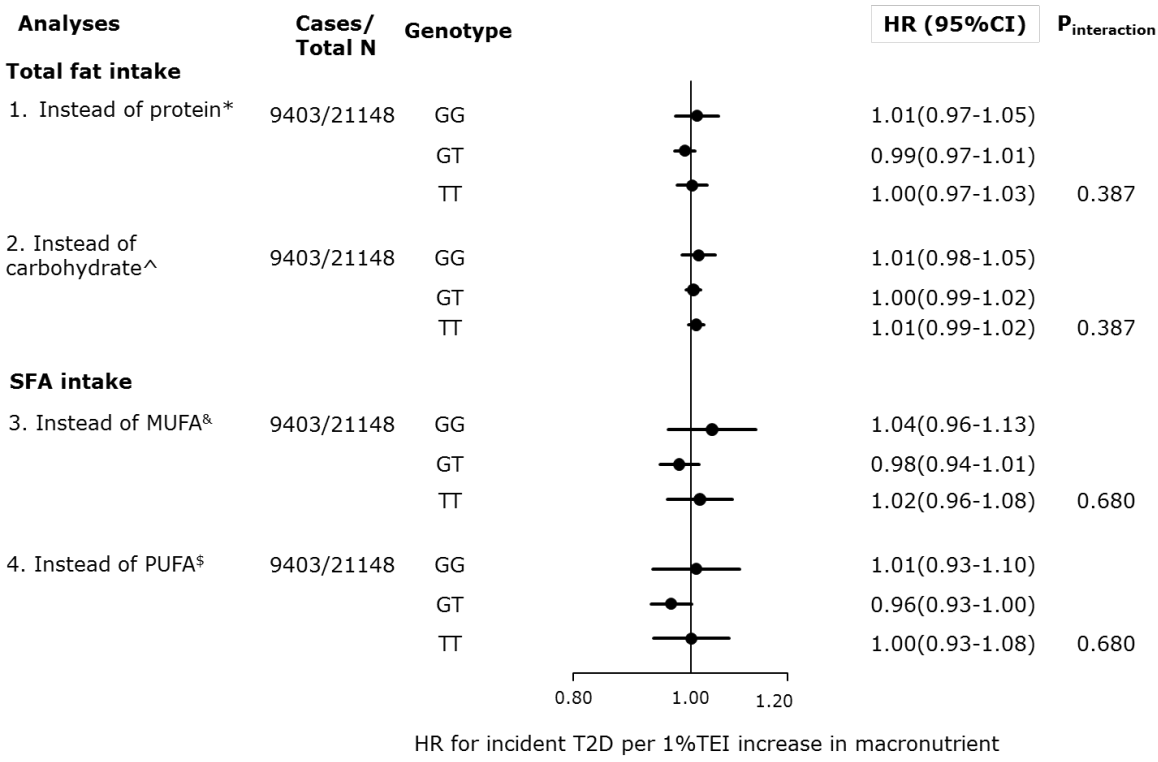



**Figure 2: EPIC-InterAct analysis for the hazard ratio (HR) of incident T2D per 1%TEI substitution in macronutrient, stratified by *CAV2* rs2270188 genotype**

Pooled HR for T2D, with adjustment for age (=underlying time scale), sex, centre, eigenvector (first 5 principal components for population stratification), physical activity, education, smoking, sex-specific alcohol categories, total energy intake (TEI), season, dietary factors (fibre, magnesium, iron, vitamin C, coffee, tea, leafy vegetables and artificially sweetened beverage) and BMI.

Models substituted for the following were further adjusted for other macronutrients * carbohydrate; ^ protein; & PUFA, protein and carbohydrate; $ MUFA, protein and carbohydrate

$P_{interaction}$: estimated by treating macronutrients and rs2270188 as continuous variables.

There was modest heterogeneity ($I^2$) between countries for total fat: 48% and SFA: 58%.

Statistical test: multiplicative interaction analysis using Prentice-weighted Cox regression

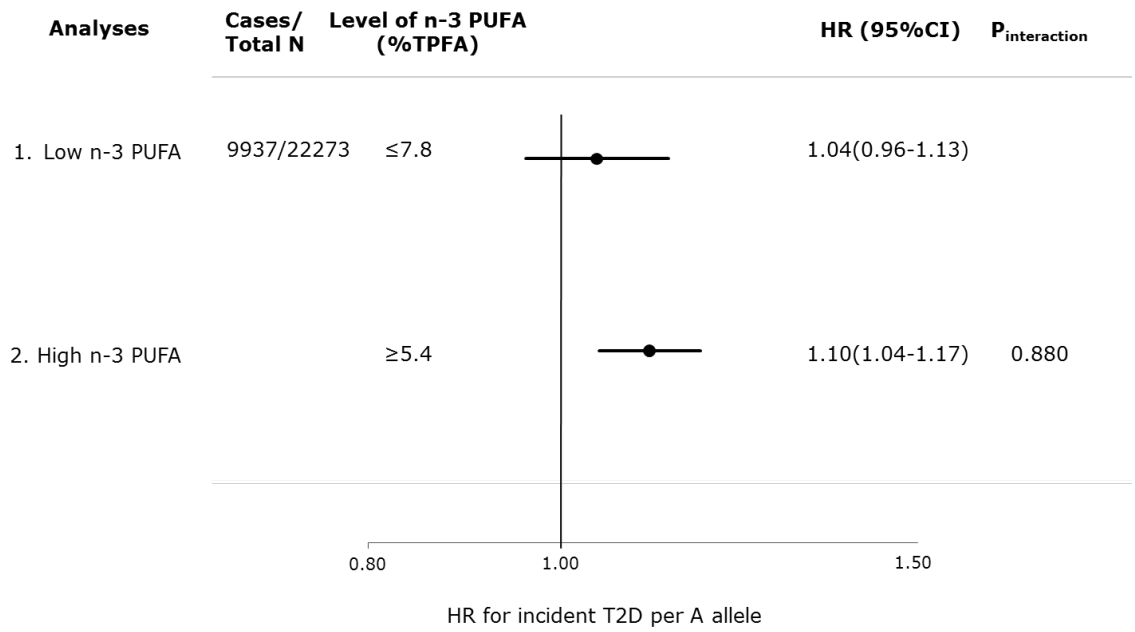



**Figure 3: EPIC-InterAct analysis for the hazard ratio (HR) of incident T2D per A allele of rs3786897 (*PEPD*), stratified by levels of circulating n-3 PUFA.**

Pooled HR for T2D, with adjustment for age (=underlying time scale), sex, centre, eigenvectors (first 5 principal components for population stratification), physical activity, education, smoking, sex-specific alcohol categories, season of blood collection, BMI

$P_{interaction}$: estimated by treating circulating n-3 PUFA and *PEPD* rs3786897 as a continuous variable.

Abbreviation: TPFA: total phospholipid fatty acid

Statistical test: multiplicative interaction analysis using Prentice-weighted Cox regression

## *Appendix E: Summary effect sizes for all studies from the systematic review of gene-macronutrient interactions and T2D*

| **Table: Summary effect sizes for all studies (if information were available)** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Gene and SNP** | **Study (n=cases/total)** | **Interaction results (G x macronutrient)*** | | | | | |
| | | **Nutrients assessed** | **Effect size** | **95%CI** | **P-value** | **P value for interaction^** | **Quality (risk of bias)** |
| ***TCF7L2*** rs7903146 | Hindy et al., 2012 (MDCS cohort n=1649/24799) | CHO | OR: 1.37 | 1.14,1.64 | 9.0x10-4 | 0.91 | Moderate<br>Large sample size with validated measurement tools and adjustment for confounders. Multiple testing not accounted for (4 tests). |
| | | Fat | OR: 1.36 | 1.13,1.64 | 9.4x10-4 | 0.47 | |
| | | PRO | OR: 1.51 | 1.29,1.77 | 3.5x10-7 | 0.70 | |
| | | Fibre | OR: 1.56 | 1.31,1.86 | 8.3x10-7 | **0.049** | |
| | | | | | | | |
| | | OR= odds per T allele at the 5$^{th}$ quintile of macronutrient intake | | | | | |
| rs7903146 rs4506565 | Wirstrom et al., 2013 (SDPP cohort n=165/5477) | Cereal fibre | | | | | Moderate<br>Exposure is not comprehensively measured and study with small sample size. |
| | | rs7903146 | OR: 1.10 | 0.76,1.59 | NA | **0.005** | |
| | | rs4506565 | OR: 1.07 | 0.74,1.54 | NA | **0.006** | |
| | | OR= odds per T allele per 5g increase of fibre/d | | | | | |
| rs12255372 | Cornelis et al., 2009 (NHS CC n=1140/3055) | CHO | OR: 1.48 | 1.20,1.82 | NA | 0.18 | Moderate<br>Well conducted study. |
| | | GI | OR: 1.54 | 1.24,1.92 | | 0.06* | |
| | | Cereal fibre | OR: 1.18 | 0.92,1.49 | | 0.14 | |
| | | GL | OR: 1.68 | 1.35,2.09 | | **0.03*** | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | After adjusting for family history (p for interaction GI: 0.14; GL: 0.13) | |
| | | OR= odds per T allele at the 3rd tertile of macronutrient intake | | | | | |
| ***GIPR*** rs10423928 | Sonestedt et al., 2012 (MDCS cohort n=1541/24840) | CHO | HR: 1.14 | 0.98,1.32 | 0.09 | **0.001** | Moderate<br>Large sample size, validated measurement tools and comprehensive analyses, with adjustment for key confounders. Multiple testing not accounted for (5 tests). |
| | | Fat | HR: 0.82 | 0.71,0.96 | 0.01 | **0.002** | |
| | | PRO | HR: 1.02 | 0.88,1.17 | | 0.83 | |
| | | Fibre | HR: 1.01 | 0.87,1.18 | | 0.22 | |
| | | Sucrose | HR: 1.02 | 0.88,1.19 | | 0.75 | |
| | | HR= hazard per A allele at the 3rd tertile of macronutrient intake | | | | | |
| ***IRS1*** rs2943641 | Ericson et al., 2013 (MDCS cohort n=1567/24841) | CHO<br>Fat<br>PRO<br>Fibre | HR: 0.80<br>HR: 0.78<br>HR: 0.99<br>HR: 0.87 | 0.61,1.08<br>0.59,1.02<br>0.75,1.30<br>0.65,1.16 | 0.75<br>0.89<br>0.65<br>0.96 | 0.59[$]<br>0.4[$]<br>0.28<br>0.92<br>[$](interaction detected for IRS1xdietxsex. P for interaction both fat and carbohydrate: 0.01) | Moderate<br>Large sample size with comprehensive dietary measurement. Multiple testing was not accounted for |
| | | HR= hazard for TT genotype with the 3rd tertile of macronutrient intake | | | | | |
| rs7578326<br>rs2943641 | Zheng et al., 2013 (GOLDN, BPRHS cross-sectional | CHO<br>Fat<br>SFA | NA | NA | | NS, NA | Moderate<br>Conducted replication and meta-analysis. Well |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | n=419/1664) | MUFA<br>SFA: CHO ratio<br>GI<br>GL | | | | | adjusted for confounding. Small sample size. |
| ***PPARG***<br>Pro12Ala/rs1801282<br><br>1431C>T | Lamri et al., 2012 (DESIR cohort n=191/4676) | Fat | NA | NA | | NS, NA | Serious<br>Dietary questionnaire with limited food items may cause measurement bias and misclassification of fat intake. Residual confounding is likely due to lack of adjustments for key confounders (total energy intake, physical activity, etc). |
| | | Ala+ | HR: 0.67 | 0.28,1.58 | | 0.05 | |
| | | ProPro | HR: 1.73 | 1.19,2.52 | | | |
| | | T+ | HR: 0.58 | 0.25,1.36 | 0.21 | | |
| | | CC | HR: 1.85 | 1.27,2.71 | 0.36 | | |
| | | HR: hazard for Ala/T allele carriers of the ProPro/CC homozygotes at the 3rd tertile of fat intake | | | | | |
| | | | | | | | |
| Pro12Ala | Cornelis et al., 2009 (NHS CC n=1140/3055) | CHO | NA | NA | | NS, NA | Moderate<br>Well conducted study. |
| | | GI | NA | NA | | NS, NA | |
| | | GL | NA | NA | | NS, NA | |
| | | Fibre | NA | NA | | NS, NA | |
| Pro12Ala | Nelson et al., 2007 (GENI study: | PUFA | NA | NA | NA | NS, NA | Serious<br>Family based association |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | family based association analysis n=736/1318) | | | | | | test conducted. Several key confounders (energy intake, BMI, physical activity, etc) not considered and possible reporting bias. |
| | | SFA | NA | NA | NA | NS, NA | |
| | | MUFA | NA | NA | NA | NS, NA | |
| | | PUFA: SFA ratio | NA | NA | NA | NS, NA | |
| Pro12Ala | Fisher et al., 2011 (EPIC-Potsdam CC 576, case-cohort 806/2864) | Fat | NA | NA | | 0.32 | Moderate<br>Robust assessment, using a novel interaction analysis approach that maximises power.<br>Multiple testing was not accounted for in exploratory analysis (256 tests). |
| 63 SNPs | | SFA | NA | NA | | 0.08 | |
| | | MUFA | NA | NA | | 0.29 | |
| | | PUFA | NA | NA | | 0.07 | |
| | | (per 1 %E increase in nutrient) | | | | | |
| ***APOA2*** −265T>C | Corella et al., 2011 (PREDIMED, SNHS cross-sectional n=825/2830) | SFA (Singapore, Asian) | OR: 3.1 | 0.87,11.02 | 0.08 | NS, NA | Moderate<br>Potential reporting bias evident. |
| | | SFA (Spain, European) | higher | NA | **0.045** | NS, NA | |
| | | OR= odds for CC genotype at the highest category of SFA intake (high vs low) | | | | | |
| ***CAV2*** | Fisher et al., 2011 | Fat | HR: 1.06 | 1.02,1.11 | **0.002** | **0.02** | Moderate |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs2270188<br><br>63 SNPs | (EPIC-Potsdam CC 576, case-cohort 2864) | | | | | | Robust assessment, using a novel interaction analysis approach that maximises power. Multiple testing was not accounted for in exploratory analysis (256 tests). |
| | | SFA | HR: 1.12 | 1.05,1.19 | **0.0006** | **0.002** | |
| | | MUFA | NA | NA | NA | | |
| | | PUFA | NA | NA | NA | | |
| | | HR= hazard for TT genotype for each 1%E increase in macronutrient intake | | | | | |
| ***FABP1/2/3/4***<br>rs2197076<br><br>12 SNPs | Mansego et al., 2012 (Hortega, Segovia replication cross-sectional n=174/2022) | Fat | OR: 1.85 | 1.03,3.84 | | 0.03274 | Serious<br>Replication conducted. The interaction was neither examined in 1/3 of the population (without reason) nor examined in SNPs without the main effect on T2D. Most confounders were not accounted for and selective reporting was evident. |
| | | SFA | NA | NA | | NA | |
| | | PUFA | | | | (not statistically significant after multiple testing corrections) | |
| | | OR= odds per A allele at the highest category of fat intake (high vs low) | | | | | |
| ***PGC-1α***<br><br>Gly482Ser<br>Thr612Met | Nelson et al., 2007 (GENI study: family-based association | PUFA | NA | NA | NA | NS, NA | Serious<br>Family-based association test conducted. Several key confounders (energy |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Thr528Thr | analysis n=736/1318) | | | | | | intake, BMI, physical activity, etc) not considered and possible reporting bias. |
| | | SFA | NA | NA | NA | NS, NA | |
| | | MUFA | NA | NA | NA | NS, NA | |
| | | PUFA: SFA ratio | NA | NA | NA | NS, NA | |
| ***PEPD*** *rs3786897* *9 SNPs examined* | Zheng et al., 2015 (case-control, n=622/915) | Circulating erythrocyte membrane phospholipid n-3 PUFA | OR: low n-3 PUFA | | | **0.027** | Serious Several likely confounders may explain the associations observed that were not adjusted for (BMI, blood lipid status, etc) and incomplete reporting of participant characteristics. |
| | | | GG vs AA: 1.88 | 1.07,3.31 | | | |
| | | | GG vs GA: 2.40 | 1.40,4.12 | | | |
| | | | high n-3 PUFA | | | | |
| | | | GG vs AA: 0.59 | 0.30,1.13 | | | |
| | | | GG vs GA : 0.65 | 0.35,1.19 | | | |
| **GRS** 15 SNPs, weighted score | Villegas et al., 2014 (NHANES cross-sectional n=1337/13120) | CHO | NA | NA | | NS, NA | Serious Dietary measurement tool was not validated and how T2D status was obtained was not described, therefore bias may be likely. The study reported a lack of statistical power to detect interactions. |
| | | Fibre | NA | NA | | NS, NA | |

Abbreviations: CHO: carbohydrate, PRO: protein, GL: glycemic load, GI: glycemic index, NS: not significant, NA: not available, MT: multiple testing, h: highest category, CC: case-control study, MDCS: Malmo Diet and Cancer Study, SDPP: Stockholm Diabetes Prevention Program, NHS: Nurse's Health Study, GOLDN: Genetics of Lipid Lowering Drugs and Diet Network, BPRHS: Boston Puerto Rican Health Study, GENI: Gene Environment Interactions study, EPIC: European Prospective Investigation into Cancer and Nutrition-Potsdam, PREDIMED: Prevención con Dieta Mediterránea trial, SNHS: , Singapore National Health Survey, NHANES: National Health and Nutrition Examination Survey.
*Interaction results: multiplicative model between gene and macronutrient. Gene is per risk allele (additive model) unless otherwise specified eg. TT genotype (codominant). Nutrient is with the highest category of macronutrient intake (categorical variable) or per X grams/%E of macronutrient (continuous variable). An example is given where there are available results.
^ p for interaction: represents the significance level for interaction across all categories of intake, under the additive genetic model, unless otherwise specified.
Gene-T2DM association: all by trait-increasing allele (additive model), except where specified otherwise.

## *Appendix F: Multiplicative models for the interaction between macronutrient and genetic risk scores (unweighted) with isocaloric macronutrient substitution: EPIC-InterAct Study*

| Macronutrient intake (5% total energy intake) | Model (*sub= substituted) | GRS for body mass index (per 6.3 risk alleles)[%] | | | GRS for insulin resistance (per 4.5 risk alleles) | | | GRS for type 2 diabetes (per 4.3 risk alleles) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **Beta (95% CI)** | ***P*** | **$I^2$ (%)** | **Beta (95% CI)** | ***P*** | **$I^2$ (%)** | **Beta (95% CI)** | ***P*** | **$I^2$ (%)** |
| **Carbohydrate** | *model 1* | 0.004(-0.024,0.032) | 0.779 | 34.3 | -0.003(-0.028,0.023) | 0.847 | 23.2 | -0.003(-0.025,0.02) | 0.815 | 1.3 |
| | *model 2* | -0.003(-0.032,0.026) | 0.849 | 34.5 | -0.009(-0.042,0.023) | 0.586 | 44.9 | -0.008(-0.031,0.015) | 0.502 | 0.0 |
| | *model 3* | -0.001(-0.033,0.032) | 0.961 | 44.5 | -0.006(-0.035,0.022) | 0.665 | 29.0 | -0.007(-0.032,0.017) | 0.555 | 5.4 |
| | *model 4* | | | | 0.004(-0.023,0.03) | 0.785 | 0.0 | 0.005(-0.029,0.039) | 0.774 | 26.5 |
| | *model 5:sub with PUFA* | -0.027(-0.141,0.087) | 0.642 | 29.2 | 0.012(-0.138,0.161) | 0.878 | 39.5 | 0.076(-0.033,0.184) | 0.172 | 0.0 |
| | *model 5:sub with MUFA* | 0.012(-0.046,0.071) | 0.680 | 0.0 | 0.033(-0.07,0.137) | 0.528 | 41.4 | -0.013(-0.108,0.083) | 0.797 | 32.5 |
| **Total protein** | *model 1* | -0.003(-0.065,0.059) | 0.923 | 11.0 | 0.05(-0.007,0.108) | 0.084 | 0.0 | -0.048(-0.112,0.015) | 0.136 | 13.7 |
| | *model 2* | 0.001(-0.079,0.081) | 0.980 | 38.1 | 0.047(-0.013,0.106) | 0.125 | 0.0 | -0.046(-0.115,0.023) | 0.194 | 18.5 |
| | *model 3* | -0.008(-0.071,0.056) | 0.815 | 4.0 | 0.056(-0.005,0.117) | 0.074 | 0.0 | -0.041(-0.123,0.04) | 0.319 | 36.9 |
| | *model 4* | | | | 0.048(-0.022,0.117) | 0.180 | 0.0 | -0.05(-0.128,0.028) | 0.213 | 15.1 |
| | *model 5:sub with carbohydrate* | 0(-0.038,0.037) | 0.991 | 57.2 | 0.005(-0.022,0.032) | 0.727 | 0.0 | 0.005(-0.028,0.037) | 0.767 | 20.9 |
| | *model 5:sub with PUFA* | -0.034(-0.147,0.08) | 0.561 | 29.2 | 0.014(-0.138,0.166) | 0.857 | 41.5 | 0.068(-0.04,0.177) | 0.215 | 0.0 |
| | *model 5:sub with MUFA* | 0.012(-0.046,0.071) | 0.676 | 0 | 0.024(-0.087,0.136) | 0.671 | 49.4 | -0.023(-0.111,0.064) | 0.601 | 23.8 |
| **Animal protein** | *model 1* | 0.009(-0.059,0.076) | 0.805 | 37.3 | 0.027(-0.025,0.08) | 0.306 | 0 | -0.041(-0.113,0.03) | 0.258 | 41.0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *model 2* | 0.007(-0.063,0.078) | 0.835 | 35.1 | 0.022(-0.033,0.077) | 0.437 | 0 | -0.036(-0.107,0.036) | 0.329 | 35.1 |
| | *model 3* | 0.005(-0.059,0.07) | 0.873 | 21.2 | 0.036(-0.031,0.103) | 0.289 | 24.4 | -0.032(-0.116,0.052) | 0.460 | 49.5 |
| | *model 4* | | | | 0.026(-0.038,0.091) | 0.425 | 0 | -0.048(-0.136,0.039) | 0.278 | 38.6 |
| | *model 5:sub with carbohydrate* | 0(-0.037,0.037) | 0.992 | 56.7 | 0.006(-0.021,0.032) | 0.687 | 0 | 0.003(-0.028,0.033) | 0.872 | 12.9 |
| | *model 5:sub with plant protein* | -0.041(-0.179,0.096) | 0.556 | 0 | 0.07(-0.106,0.247) | 0.434 | 12.5 | 0(-0.18,0.179) | 0.996 | 12.3 |
| **Plant protein** | *model 1* | -0.041(-0.168,0.087) | 0.533 | 0.0 | 0.059(-0.112,0.231) | 0.499 | 32.8 | 0.025(-0.154,0.205) | 0.780 | 40.0 |
| | *model 2* | -0.059(-0.191,0.073) | 0.384 | 0 | 0.066(-0.11,0.242) | 0.463 | 31.5 | 0.01(-0.176,0.196) | 0.917 | 40.6 |
| | *model 3* | -0.06(-0.197,0.076) | 0.384 | 0.0 | 0.094(-0.07,0.259) | 0.260 | 19.3 | 0.01(-0.171,0.191) | 0.916 | 34.3 |
| | *model 4* | | | | 0.079(-0.084,0.243) | 0.342 | 3.6 | -0.009(-0.192,0.173) | 0.921 | 15.6 |
| | *model 5:sub with carbohydrate* | 0.001(-0.037,0.038) | 0.974 | 57.2 | 0.002(-0.026,0.031) | 0.869 | 7.6 | 0.003(-0.028,0.034) | 0.855 | 14.7 |
| **Total fat** | *model 1* | 0.008(-0.017,0.033) | 0.552 | 0.0 | -0.005(-0.037,0.028) | 0.772 | 32.0 | 0.017(-0.009,0.044) | 0.188 | 0.0 |
| | *model 2* | 0.012(-0.014,0.037) | 0.375 | 0.0 | -0.004(-0.041,0.033) | 0.833 | 41.7 | 0.016(-0.011,0.043) | 0.248 | 0.0 |
| | *model 3* | 0.01(-0.017,0.036) | 0.471 | 0.0 | -0.007(-0.04,0.025) | 0.653 | 22.5 | 0.016(-0.012,0.045) | 0.267 | 5.9 |
| | *model 4* | | | | -0.004(-0.048,0.041) | 0.869 | 41.2 | 0.011(-0.021,0.043) | 0.497 | 0.0 |
| | *model 5:sub with carbohydrate* | 0.001(-0.032,0.034) | 0.955 | 44.3 | 0(-0.028,0.029) | 0.986 | 9.3 | 0.004(-0.03,0.037) | 0.833 | 23.6 |
| | *model 5:sub with total protein* | -0.008(-0.074,0.059) | 0.821 | 10.9 | 0.049(-0.021,0.119) | 0.168 | 0 | -0.057(-0.138,0.024) | 0.167 | 19.7 |
| | *model 5:sub with plant protein* | -0.058(-0.197,0.08) | 0.407 | 0 | 0.066(-0.116,0.248) | 0.475 | 16.1 | -0.014(-0.21,0.183) | 0.892 | 24.0 |
| | *model 5:sub with animal protein* | 0.004(-0.061,0.069) | 0.898 | 21.2 | 0.034(-0.035,0.103) | 0.335 | 9.1 | -0.05(-0.134,0.034) | 0.247 | 33.3 |
| **SFA** | *model 1* | 0.022(-0.03,0.073) | 0.407 | 12.2 | -0.019(-0.068,0.03) | 0.444 | 0 | 0.047(-0.003,0.096) | 0.066 | 0 |
| | *model 2* | 0.033(-0.018,0.085) | 0.208 | 8.5 | -0.021(-0.073,0.03) | 0.412 | 0 | 0.045(-0.006,0.096) | 0.085 | 0 |
| | *model 3* | 0.031(-0.019,0.08) | 0.228 | 0 | -0.027(-0.08,0.025) | 0.311 | 0 | 0.046(-0.007,0.1) | 0.09 | 2.9 |
| | *model 4* | | | | -0.015(-0.089,0.058) | 0.687 | 22.8 | 0.027(-0.034,0.089) | 0.384 | 0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *model 5:sub with carbohydrate* | 0.001(-0.036,0.038) | 0.957 | 54.4 | 0.003(-0.024,0.03) | 0.85 | 0 | 0.001(-0.031,0.033) | 0.941 | 18.7 |
| | *model 5:sub with plant protein* | -0.045(-0.183,0.094) | 0.526 | 0 | 0.064(-0.114,0.243) | 0.479 | 13.4 | -0.004(-0.185,0.178) | 0.966 | 13.7 |
| | *model 5:sub with animal protein* | 0.002(-0.061,0.064) | 0.96 | 14.1 | 0.027(-0.039,0.092) | 0.425 | 0 | -0.052(-0.14,0.037) | 0.25 | 38.8 |
| | *model 5:sub with total protein* | -0.01(-0.072,0.052) | 0.753 | 0 | 0.046(-0.025,0.116) | 0.206 | 0 | -0.054(-0.135,0.028) | 0.195 | 19.7 |
| | *model 5:sub with PUFA* | -0.031(-0.146,0.083) | 0.593 | 29.6 | 0.006(-0.142,0.153) | 0.939 | 37.9 | 0.075(-0.033,0.183) | 0.173 | 0 |
| | *model 5:sub with MUFA* | 0.013(-0.046,0.072) | 0.671 | 0 | 0.035(-0.073,0.143) | 0.528 | 45.3 | -0.008(-0.105,0.089) | 0.869 | 33.7 |
| **MUFA** | *model 1* | 0.01(-0.046,0.066) | 0.724 | 0 | -0.001(-0.079,0.077) | 0.982 | 35.7 | 0.002(-0.075,0.08) | 0.95 | 34.0 |
| | *model 2* | 0.015(-0.043,0.072) | 0.617 | 0 | -0.005(-0.083,0.073) | 0.902 | 31.7 | 0.001(-0.083,0.084) | 0.987 | 38.9 |
| | *model 3* | 0.014(-0.044,0.073) | 0.634 | 0 | -0.003(-0.088,0.083) | 0.949 | 38.5 | -0.003(-0.092,0.087) | 0.954 | 43.8 |
| | *model 4* | | | | 0.023(-0.087,0.133) | 0.685 | 48.5 | -0.016(-0.097,0.065) | 0.699 | 15.3 |
| | *model 5:sub with plant protein* | -0.045(-0.183,0.094) | 0.527 | 0 | 0.054(-0.137,0.246) | 0.576 | 21.9 | -0.01(-0.197,0.177) | 0.916 | 17.5 |
| | *model 5:sub with animal protein* | 0(-0.066,0.066) | 0.996 | 22.4 | 0.031(-0.036,0.098) | 0.37 | 3.1 | -0.052(-0.136,0.032) | 0.226 | 33.2 |
| **PUFA** | *model 1* | -0.042(-0.124,0.039) | 0.309 | 0 | -0.019(-0.129,0.092) | 0.742 | 32.9 | 0.036(-0.05,0.123) | 0.408 | 0 |
| | *model 2* | -0.039(-0.141,0.064) | 0.459 | 21.9 | -0.025(-0.119,0.068) | 0.598 | 7.97 | 0.044(-0.047,0.134) | 0.342 | 0 |
| | *model 3* | -0.042(-0.155,0.072) | 0.471 | 30.5 | -0.019(-0.109,0.071) | 0.676 | 1.1 | 0.044(-0.048,0.136) | 0.351 | 0 |
| | *model 4* | | | | 0.01(-0.139,0.16) | 0.895 | 40.4 | 0.072(-0.039,0.182) | 0.204 | 2.2 |
| | *model 5:sub with plant protein* | -0.046(-0.185,0.092) | 0.51 | 0 | 0.065(-0.111,0.241) | 0.467 | 11.5 | -0.004(-0.198,0.189) | 0.964 | 22.1 |
| | *model 5:sub with animal protein* | 0.002(-0.059,0.064) | 0.938 | 12.2 | 0.028(-0.038,0.093) | 0.41 | 0 | -0.05(-0.135,0.036) | 0.255 | 35.3 |
| | *model 5:sub with MUFA* | 0.018(-0.04,0.077) | 0.544 | 0 | 0.034(-0.076,0.144) | 0.545 | 47.5 | -0.01(-0.109,0.089) | 0.842 | 36.2 |
| **Total dietary** | *model 1* | 0.005(-0.005,0.015) | 0.346 | 0 | 0(-0.011,0.012) | 0.95 | 20.9 | 0.001(-0.009,0.011) | 0.86 | 0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **fibre** | | | | | | | | | | |
| | *model 2* | 0.005(-0.006,0.015) | 0.382 | 0 | -0.003(-0.017,0.012) | 0.723 | 36.5 | 0.001(-0.01,0.011) | 0.919 | 0 |
| | *model 3* | 0.005(-0.006,0.015) | 0.385 | 0 | -0.004(-0.017,0.01) | 0.602 | 28.6 | 0.002(-0.009,0.013) | 0.76 | 0 |
| | *model 4* | | | | 0(-0.015,0.016) | 0.951 | 24.5 | 0.001(-0.011,0.013) | 0.875 | 0 |
| **Vegetable fibre** | *model 1* | -0.001(-0.026,0.024) | 0.921 | 0 | -0.008(-0.045,0.03) | 0.688 | 44.7 | 0.003(-0.023,0.03) | 0.81 | 0 |
| | *model 2* | 0.004(-0.026,0.033) | 0.802 | 11.4 | -0.013(-0.054,0.029) | 0.553 | 51.1 | 0.002(-0.025,0.029) | 0.897 | 0 |
| | *model 3* | -0.002(-0.033,0.029) | 0.903 | 10.6 | -0.009(-0.048,0.031) | 0.66 | 41.1 | 0.007(-0.021,0.035) | 0.618 | 0 |
| | *model 4* | | | | -0.011(-0.042,0.02) | 0.49 | 2.1 | 0.013(-0.019,0.044) | 0.433 | 0 |
| **Fruit fibre** | *model 1* | 0.006(-0.012,0.024) | 0.53 | 0 | 0(-0.026,0.027) | 0.981 | 40.1 | -0.003(-0.029,0.022) | 0.79 | 33.7 |
| | *model 2* | 0.005(-0.016,0.027) | 0.637 | 16.2 | -0.003(-0.031,0.025) | 0.83 | 43.1 | -0.004(-0.034,0.026) | 0.799 | 45.4 |
| | *model 3* | 0.004(-0.015,0.023) | 0.662 | 0 | -0.006(-0.036,0.025) | 0.721 | 47.1 | -0.002(-0.036,0.033) | 0.928 | 54.5 |
| | *model 4* | | | | 0.001(-0.042,0.045) | 0.948 | 66.5 | -0.001(-0.043,0.041) | 0.965 | 60.2 |
| **Cereal fibre** | *model 1* | -0.003(-0.019,0.013) | 0.695 | 0 | 0.003(-0.014,0.02) | 0.742 | 0 | 0.002(-0.017,0.022) | 0.829 | 17.0 |
| | *model 2* | -0.006(-0.023,0.01) | 0.452 | 0 | -0.001(-0.018,0.017) | 0.925 | 0 | 0.006(-0.012,0.024) | 0.52 | 5.5 |
| | *model 3* | -0.006(-0.023,0.011) | 0.472 | 0 | -0.002(-0.021,0.016) | 0.795 | 4.2 | 0.006(-0.011,0.024) | 0.476 | 0 |
| | *model 4* | | | | -0.008(-0.029,0.014) | 0.489 | 5.5 | 0.006(-0.016,0.028) | 0.606 | 9.8 |

Abbreviations: SFA- saturated fatty acid, MUFA- monounsaturated fatty acid, PUFA- polyunsaturated fatty acid, GRS- genetic risk score

Macronutrients are represented by per 5% of total energy intake and dietary fibre by per g/1000kcal.

**Beta-coefficient for the interaction between each of the genetic risk score with the following macronutrients on incident T2D are adjusted for the following covariates.**

**Macronutrient**

Model 1: age (=underlying time scale), sex, centre, total energy (TEI), first 5 principal component (PC) for population stratification

Model 2: model 1 + lifestyle factors- physical activity, education, smoking, sex-specific alcohol categories

Model 3: model 2+ dietary covariates (dietary fibre, magnesium, iron, vitamin C, leafy vegetables, tea, coffee)

Model 4: model 3+ BMI

Model 5: model 4+ isocaloric macronutrient substitution

**Dietary fibre**

Model 1: age (=underlying time scale), sex, centre, TEI, first 5 PC for population stratification

Model 2: model 1 + lifestyle factors- physical activity, education, smoking, sex-specific alcohol categories

Model 3: model 2+ dietary covariates (carbohydrate, SFA, MUFA, PUFA intake, magnesium, iron, vitamin C, leafy vegetable, tea, coffee )

Model 4: model 3+ BMI

fibre subtypes: last model includes mutual adjustment

[&] There is no adjustment for BMI for interactions models with BMI GRS

**Example of interpretation:** the beta-coefficient of the interaction between total fat and BMI GRS was 0.001 for incident T2D, when fat replaced carbohydrate intake. However, this was not statistically significant because the 95% confidence interval is -0.032 and 0.034.

**Isocaloric macronutrient substitution: α for significant interaction< 6.17E-4 (0.05/81 tests)**

## *Appendix G: Individual SNP and macronutrient interactions and incident T2D, with p-value for interaction<0.05*

Please note that these interactions are based on isocaloric macronutrient substitution modelling detailed in Appendix F.

GRS: genetic risk score
snp: single nucleotide polymorphism
macro: macronutrient intake
submacro: macronutrient being substituted for (this is being replaced)
beta: beta coefficient
se: standard error
lci: lower confidence interval
uci: upper confidence interval
p_int: pvalue for interaction
i_sq: I squared for heterogeneity
p_het: pvalue for heterogeneity

Threshold for p value for interaction after Bonferroni correction< 9.4E-06

| GRS | snp | macro | submac | beta | se | lci | uci | p_int | i_sq | p_het |
|---|---|---|---|---|---|---|---|---|---|---|
| T2D | rs3802177 | mufa5Ew | pufa5Ew | -0.194 | 0.053 | -0.298 | -0.089 | 2.954E-04 | 0.000 | 0.948 |
| T2D | rs3802177 | mufa5Ew | prot5Ew | -0.189 | 0.053 | -0.293 | -0.085 | 3.782E-04 | 0.000 | 0.877 |
| T2D | rs3802177 | mufa5Ew | sfa5Ew | -0.188 | 0.053 | -0.293 | -0.084 | 4.283E-04 | 0.000 | 0.795 |
| BMI | rs17724992 | prot5Ew | fat5Ew | -0.179 | 0.051 | -0.279 | -0.079 | 4.473E-04 | 0.000 | 0.533 |
| BMI | rs17724992 | prot5Ew | sfa5Ew | -0.178 | 0.051 | -0.278 | -0.077 | 0.001 | 0.000 | 0.543 |
| T2D | rs3802177 | mufa5Ew | cho5Ew | -0.184 | 0.053 | -0.288 | -0.080 | 0.001 | 0.000 | 0.653 |
| IR | rs6937438 | pufa5Ew | prot5Ew | -0.279 | 0.084 | -0.443 | -0.115 | 0.001 | 0.000 | 0.469 |
| IR | rs6937438 | pufa5Ew | sfa5Ew | -0.277 | 0.084 | -0.441 | -0.112 | 0.001 | 0.000 | 0.437 |
| IR | rs6937438 | pufa5Ew | cho5Ew | -0.274 | 0.084 | -0.439 | -0.110 | 0.001 | 0.000 | 0.461 |
| BMI | rs3849570 | cho5Ew | sfa5Ew | -0.057 | 0.018 | -0.092 | -0.023 | 0.001 | 0.000 | 0.579 |
| BMI | rs3849570 | cho5Ew | fat5Ew | -0.057 | 0.018 | -0.091 | -0.022 | 0.001 | 0.000 | 0.696 |
| BMI | rs3849570 | cho5Ew | protp5Ew | -0.056 | 0.018 | -0.091 | -0.022 | 0.001 | 0.000 | 0.580 |
| BMI | rs3849570 | cho5Ew | prot5Ew | -0.055 | 0.017 | -0.089 | -0.021 | 0.001 | 0.000 | 0.658 |
| BMI | rs3849570 | cho5Ew | prota5Ew | -0.055 | 0.017 | -0.089 | -0.021 | 0.002 | 0.000 | 0.676 |
| BMI | rs1928295 | prot5Ew | sfa5Ew | 0.133 | 0.044 | 0.046 | 0.221 | 0.003 | 0.000 | 0.877 |
| BMI | rs1928295 | prot5Ew | fat5Ew | 0.130 | 0.044 | 0.044 | 0.217 | 0.003 | 0.000 | 0.794 |
| IR | rs2249105 | fb_fruit_k | fruit | 0.044 | 0.016 | 0.013 | 0.075 | 0.005 | 0.000 | 0.816 |
| T2D | rs6878122 | pufa5Ew | sfa5Ew | -0.239 | 0.089 | -0.413 | -0.066 | 0.007 | 8.402 | 0.365 |
| BMI | rs4256980 | prota5Ew | pufa5Ew | -0.110 | 0.041 | -0.191 | -0.029 | 0.008 | 0.000 | 0.691 |
| BMI | rs17724992 | prota5Ew | mufa5Ew | -0.140 | 0.053 | -0.243 | -0.037 | 0.008 | 17.576 | 0.291 |
| BMI | rs9540493 | protp5Ew | prota5Ew | -0.269 | 0.101 | -0.467 | -0.071 | 0.008 | 0.000 | 0.882 |
| T2D | rs6878122 | pufa5Ew | cho5Ew | -0.246 | 0.093 | -0.428 | -0.064 | 0.008 | 13.219 | 0.327 |
| BMI | rs17724992 | prota5Ew | sfa5Ew | -0.142 | 0.054 | -0.248 | -0.037 | 0.008 | 20.478 | 0.267 |
| BMI | rs1928295 | prota5Ew | mufa5Ew | 0.106 | 0.040 | 0.027 | 0.185 | 0.009 | 0.000 | 0.652 |
| BMI | rs1928295 | prota5Ew | pufa5Ew | 0.106 | 0.040 | 0.027 | 0.185 | 0.009 | 0.000 | 0.592 |
| IR | rs17402950 | fb_cereal_ | fruit | -0.091 | 0.035 | -0.160 | -0.023 | 0.009 | 0.000 | 0.537 |
| BMI | rs4256980 | prota5Ew | fat5Ew | -0.108 | 0.041 | -0.189 | -0.027 | 0.009 | 0.000 | 0.751 |
| T2D | rs6878122 | pufa5Ew | prot5Ew | -0.265 | 0.102 | -0.466 | -0.065 | 0.009 | 24.812 | 0.231 |
| BMI | rs17724992 | prota5Ew | pufa5Ew | -0.136 | 0.053 | -0.240 | -0.033 | 0.010 | 18.324 | 0.285 |
| BMI | rs1928295 | prota5Ew | sfa5Ew | 0.104 | 0.040 | 0.024 | 0.183 | 0.010 | 0.000 | 0.683 |
| BMI | rs17724992 | prota5Ew | fat5Ew | -0.143 | 0.056 | -0.253 | -0.033 | 0.011 | 26.509 | 0.217 |
| BMI | rs3101336 | cho5Ew | prota5Ew | 0.043 | 0.017 | 0.010 | 0.077 | 0.011 | 0.000 | 0.675 |
| BMI | rs3101336 | cho5Ew | prot5Ew | 0.043 | 0.017 | 0.010 | 0.076 | 0.011 | 0.000 | 0.623 |
| BMI | rs4256980 | prot5Ew | fat5Ew | -0.115 | 0.045 | -0.204 | -0.026 | 0.011 | 0.000 | 0.744 |
| BMI | rs12446632 | fb_kcalw | fruitw | 0.027 | 0.011 | 0.006 | 0.048 | 0.011 | 0.000 | 0.707 |
| BMI | rs3101336 | cho5Ew | protp5Ew | 0.043 | 0.017 | 0.010 | 0.077 | 0.011 | 0.000 | 0.658 |
| BMI | rs4256980 | prota5Ew | mufa5Ew | -0.104 | 0.042 | -0.186 | -0.023 | 0.012 | 0.000 | 0.652 |
| BMI | rs9540493 | protp5Ew | sfa5Ew | -0.255 | 0.102 | -0.454 | -0.056 | 0.012 | 0.000 | 0.779 |
| BMI | rs9540493 | protp5Ew | pufa5Ew | -0.254 | 0.102 | -0.453 | -0.055 | 0.012 | 0.000 | 0.707 |
| BMI | rs6091540 | fb_veg_kc | fruit | 0.066 | 0.026 | 0.014 | 0.118 | 0.012 | 19.607 | 0.274 |
| BMI | rs1928295 | prota5Ew | fat5Ew | 0.101 | 0.040 | 0.022 | 0.180 | 0.012 | 0.885 | 0.422 |
| BMI | rs7899106 | pufa5Ew | prot5Ew | 0.336 | 0.135 | 0.072 | 0.600 | 0.013 | 0.000 | 0.815 |
| BMI | rs3101336 | cho5Ew | sfa5Ew | 0.043 | 0.017 | 0.009 | 0.077 | 0.013 | 0.000 | 0.779 |
| BMI | rs7164727 | fb_kcalw | fruitw | 0.020 | 0.008 | 0.004 | 0.036 | 0.013 | 0.000 | 0.899 |
| T2D | rs780094 | mufa5Ew | prot5Ew | -0.122 | 0.049 | -0.219 | -0.026 | 0.013 | 0.000 | 0.435 |
| BMI | rs9540493 | protp5Ew | fat5Ew | -0.250 | 0.101 | -0.449 | -0.052 | 0.013 | 0.000 | 0.681 |
| BMI | rs4256980 | prot5Ew | sfa5Ew | -0.112 | 0.046 | -0.202 | -0.022 | 0.014 | 0.000 | 0.669 |
| T2D | rs2943640 | prot5Ew | sfa5Ew | 0.134 | 0.055 | 0.026 | 0.241 | 0.015 | 3.893 | 0.400 |
| BMI | rs9540493 | protp5Ew | mufa5Ew | -0.248 | 0.102 | -0.448 | -0.049 | 0.015 | 0.000 | 0.822 |
| BMI | rs2112347 | fb_cereal_ | fruit | -0.035 | 0.014 | -0.063 | -0.007 | 0.015 | 10.756 | 0.347 |
| BMI | rs4256980 | prota5Ew | sfa5Ew | -0.101 | 0.042 | -0.182 | -0.019 | 0.016 | 0.000 | 0.728 |
| BMI | rs3101336 | cho5Ew | fat5Ew | 0.041 | 0.017 | 0.008 | 0.075 | 0.016 | 0.000 | 0.751 |
| T2D | rs163184 | fb_kcalw | fruitw | 0.022 | 0.009 | 0.004 | 0.041 | 0.017 | 7.170 | 0.375 |
| T2D | rs12427353 | fb_cereal_ | fruit | -0.048 | 0.020 | -0.088 | -0.008 | 0.019 | 12.708 | 0.331 |
| BMI | rs7715256 | prot5Ew | fat5Ew | -0.102 | 0.044 | -0.189 | -0.015 | 0.021 | 0.000 | 0.922 |
| BMI | rs11126666 | cho5Ew | fat5Ew | 0.044 | 0.019 | 0.006 | 0.081 | 0.022 | 0.000 | 0.600 |
| IR | rs308971 | prot5Ew | fat5Ew | -0.225 | 0.098 | -0.418 | -0.032 | 0.022 | 32.020 | 0.172 |

| GRS | snp | macro | submac | beta | se | lci | uci | p_int | i_sq | p_het |
|---|---|---|---|---|---|---|---|---|---|---|
| BMI | rs12885454 | mufa5Ew | prot5Ew | -0.100 | 0.044 | -0.185 | -0.014 | 0.022 | 0.000 | 0.630 |
| BMI | rs1167827 | fb_veg_kc | fruit | -0.060 | 0.026 | -0.111 | -0.008 | 0.023 | 30.928 | 0.181 |
| BMI | rs7899106 | pufa5Ew | cho5Ew | 0.310 | 0.136 | 0.043 | 0.577 | 0.023 | 0.000 | 0.780 |
| BMI | rs7715256 | prot5Ew | sfa5Ew | -0.100 | 0.045 | -0.187 | -0.012 | 0.025 | 0.000 | 0.757 |
| T2D | rs2943640 | prot5Ew | fat5Ew | 0.132 | 0.059 | 0.016 | 0.248 | 0.026 | 15.231 | 0.310 |
| BMI | rs492400 | fb_kcalw | fruitw | -0.018 | 0.008 | -0.033 | -0.002 | 0.026 | 0.000 | 0.638 |
| IR | rs308971 | prot5Ew | sfa5Ew | -0.221 | 0.099 | -0.416 | -0.026 | 0.026 | 31.916 | 0.173 |
| BMI | rs10132280 | fb_fruit_k | fruit | -0.035 | 0.016 | -0.066 | -0.004 | 0.026 | 6.174 | 0.383 |
| BMI | rs9540493 | cho5Ew | prota5Ew | -0.040 | 0.018 | -0.075 | -0.005 | 0.026 | 8.880 | 0.361 |
| IR | rs731839 | protp5Ew | prota5Ew | -0.268 | 0.122 | -0.507 | -0.029 | 0.028 | 0.000 | 0.837 |
| IR | rs11130329 | protp5Ew | prota5Ew | 0.437 | 0.199 | 0.047 | 0.828 | 0.028 | 27.652 | 0.208 |
| BMI | rs11057405 | mufa5Ew | pufa5Ew | -0.180 | 0.082 | -0.341 | -0.019 | 0.029 | 0.000 | 0.701 |
| BMI | rs2365389 | pufa5Ew | prot5Ew | 0.141 | 0.065 | 0.015 | 0.268 | 0.029 | 0.000 | 0.806 |
| BMI | rs9540493 | cho5Ew | prot5Ew | -0.041 | 0.019 | -0.078 | -0.004 | 0.029 | 15.152 | 0.311 |
| IR | rs7227237 | pufa5Ew | prot5Ew | 0.193 | 0.089 | 0.019 | 0.366 | 0.029 | 0.000 | 0.698 |
| IR | rs11130329 | protp5Ew | mufa5Ew | 0.401 | 0.185 | 0.038 | 0.763 | 0.030 | 18.978 | 0.280 |
| T2D | rs10203174 | fb_fruit_k | fruit | 0.096 | 0.044 | 0.009 | 0.182 | 0.030 | 56.636 | 0.024 |
| BMI | rs1516725 | mufa5Ew | prot5Ew | -0.137 | 0.063 | -0.261 | -0.013 | 0.031 | 0.000 | 0.724 |
| IR | rs2943645 | prot5Ew | sfa5Ew | 0.127 | 0.059 | 0.012 | 0.242 | 0.031 | 14.153 | 0.319 |
| IR | rs11130329 | protp5Ew | pufa5Ew | 0.430 | 0.199 | 0.039 | 0.821 | 0.031 | 27.585 | 0.208 |
| IR | rs308971 | prota5Ew | fat5Ew | -0.215 | 0.100 | -0.411 | -0.020 | 0.031 | 43.961 | 0.086 |
| BMI | rs2112347 | protp5Ew | prota5Ew | -0.216 | 0.100 | -0.413 | -0.020 | 0.031 | 0.000 | 0.778 |
| BMI | rs11057405 | mufa5Ew | sfa5Ew | -0.178 | 0.083 | -0.340 | -0.016 | 0.031 | 0.000 | 0.697 |
| BMI | rs7899106 | pufa5Ew | sfa5Ew | 0.297 | 0.138 | 0.026 | 0.567 | 0.032 | 0.000 | 0.734 |
| BMI | rs11126666 | cho5Ew | protp5Ew | 0.041 | 0.019 | 0.004 | 0.078 | 0.032 | 0.000 | 0.702 |
| BMI | rs3736485 | cho5Ew | prota5Ew | -0.037 | 0.017 | -0.071 | -0.003 | 0.032 | 3.069 | 0.406 |
| T2D | rs459193 | pufa5Ew | prot5Ew | -0.173 | 0.081 | -0.332 | -0.014 | 0.032 | 0.000 | 0.659 |
| IR | rs459193 | pufa5Ew | prot5Ew | -0.173 | 0.081 | -0.332 | -0.014 | 0.032 | 0.000 | 0.659 |
| BMI | rs758747 | prota5Ew | mufa5Ew | -0.097 | 0.045 | -0.186 | -0.008 | 0.033 | 0.000 | 0.453 |
| IR | rs7227237 | pufa5Ew | sfa5Ew | 0.189 | 0.089 | 0.015 | 0.363 | 0.033 | 0.000 | 0.716 |
| BMI | rs2112347 | fb_veg_kc | fruit | -0.045 | 0.021 | -0.087 | -0.004 | 0.033 | 4.548 | 0.395 |
| BMI | rs11126666 | cho5Ew | sfa5Ew | 0.041 | 0.019 | 0.003 | 0.079 | 0.033 | 0.000 | 0.685 |
| IR | rs9492443 | fb_veg_kc | fruit | 0.055 | 0.026 | 0.004 | 0.105 | 0.033 | 0.000 | 0.879 |
| BMI | rs2365389 | pufa5Ew | cho5Ew | 0.138 | 0.065 | 0.011 | 0.265 | 0.033 | 0.000 | 0.859 |
| BMI | rs1167827 | fb_fruit_k | fruit | -0.029 | 0.014 | -0.056 | -0.002 | 0.033 | 0.000 | 0.589 |
| IR | rs11130329 | protp5Ew | sfa5Ew | 0.407 | 0.192 | 0.032 | 0.783 | 0.034 | 22.975 | 0.246 |
| BMI | rs2033529 | fb_kcalw | fruitw | 0.018 | 0.008 | 0.001 | 0.034 | 0.034 | 0.000 | 0.851 |
| BMI | rs7239883 | prot5Ew | sfa5Ew | 0.096 | 0.045 | 0.007 | 0.185 | 0.034 | 0.000 | 0.795 |
| BMI | rs12885454 | mufa5Ew | pufa5Ew | -0.093 | 0.044 | -0.180 | -0.007 | 0.034 | 0.000 | 0.532 |
| BMI | rs7164727 | fb_cereal_ | fruit | 0.030 | 0.014 | 0.002 | 0.058 | 0.035 | 6.473 | 0.380 |
| BMI | rs758747 | prota5Ew | fat5Ew | -0.095 | 0.045 | -0.184 | -0.007 | 0.035 | 0.000 | 0.507 |
| BMI | rs2033732 | protp5Ew | fat5Ew | -0.252 | 0.120 | -0.487 | -0.018 | 0.035 | 0.000 | 0.495 |
| IR | rs308971 | prota5Ew | mufa5Ew | -0.216 | 0.102 | -0.417 | -0.015 | 0.035 | 45.999 | 0.073 |
| T2D | rs2943640 | protp5Ew | pufa5Ew | 0.251 | 0.119 | 0.017 | 0.485 | 0.035 | 0.000 | 0.534 |
| BMI | rs11126666 | cho5Ew | prot5Ew | 0.040 | 0.019 | 0.003 | 0.076 | 0.036 | 0.000 | 0.578 |
| BMI | rs13107325 | prota5Ew | pufa5Ew | -0.168 | 0.080 | -0.325 | -0.011 | 0.036 | 0.000 | 0.612 |
| BMI | rs2112347 | protp5Ew | mufa5Ew | -0.212 | 0.101 | -0.410 | -0.014 | 0.036 | 0.000 | 0.676 |
| IR | rs7227237 | pufa5Ew | cho5Ew | 0.186 | 0.088 | 0.012 | 0.359 | 0.036 | 0.000 | 0.715 |
| T2D | rs780094 | mufa5Ew | pufa5Ew | -0.111 | 0.053 | -0.215 | -0.007 | 0.036 | 5.329 | 0.389 |
| IR | rs731839 | protp5Ew | sfa5Ew | -0.257 | 0.122 | -0.497 | -0.017 | 0.036 | 0.000 | 0.833 |
| BMI | rs11126666 | cho5Ew | prota5Ew | 0.039 | 0.019 | 0.003 | 0.076 | 0.036 | 0.000 | 0.601 |
| T2D | rs2943640 | protp5Ew | prota5Ew | 0.248 | 0.119 | 0.016 | 0.481 | 0.036 | 0.000 | 0.611 |
| BMI | rs11057405 | mufa5Ew | prot5Ew | -0.170 | 0.081 | -0.330 | -0.010 | 0.037 | 0.000 | 0.719 |
| T2D | rs780094 | mufa5Ew | sfa5Ew | -0.111 | 0.053 | -0.214 | -0.007 | 0.037 | 5.075 | 0.391 |
| T2D | rs2943640 | protp5Ew | fat5Ew | 0.248 | 0.119 | 0.015 | 0.481 | 0.037 | 0.000 | 0.572 |
| T2D | rs3802177 | protp5Ew | pufa5Ew | 0.263 | 0.127 | 0.015 | 0.511 | 0.037 | 0.000 | 0.968 |
| BMI | rs2365389 | pufa5Ew | sfa5Ew | 0.135 | 0.065 | 0.008 | 0.262 | 0.038 | 0.000 | 0.814 |
| T2D | rs11634397 | pufa5Ew | prot5Ew | -0.164 | 0.079 | -0.320 | -0.009 | 0.038 | 0.000 | 0.867 |
| BMI | rs7239883 | prot5Ew | fat5Ew | 0.094 | 0.045 | 0.005 | 0.182 | 0.038 | 0.000 | 0.779 |
| BMI | rs6091540 | fb_kcalw | fruitw | 0.018 | 0.009 | 0.001 | 0.035 | 0.038 | 0.000 | 0.591 |
| IR | rs308971 | prota5Ew | sfa5Ew | -0.203 | 0.098 | -0.396 | -0.011 | 0.038 | 40.995 | 0.105 |

| GRS | snp | macro | submac | beta | se | lci | uci | p_int | i_sq | p_het |
|---|---|---|---|---|---|---|---|---|---|---|
| BMI | rs10733682 | cho5Ew | prot5Ew | 0.039 | 0.019 | 0.002 | 0.076 | 0.038 | 15.983 | 0.304 |
| IR | rs11130329 | protp5Ew | fat5Ew | 0.406 | 0.196 | 0.022 | 0.790 | 0.038 | 26.203 | 0.220 |
| IR | rs3864041 | fb_cereal_ | fruit | 0.044 | 0.021 | 0.002 | 0.085 | 0.038 | 37.562 | 0.130 |
| BMI | rs2033732 | protp5Ew | mufa5Ew | -0.249 | 0.120 | -0.485 | -0.013 | 0.038 | 0.000 | 0.633 |
| T2D | rs2075423 | fb_cereal_ | fruit | 0.035 | 0.017 | 0.002 | 0.068 | 0.039 | 8.408 | 0.365 |
| BMI | rs1516725 | mufa5Ew | pufa5Ew | -0.132 | 0.064 | -0.258 | -0.007 | 0.039 | 0.000 | 0.873 |
| BMI | rs11057405 | mufa5Ew | cho5Ew | -0.170 | 0.082 | -0.331 | -0.008 | 0.039 | 0.000 | 0.721 |
| BMI | rs2112347 | protp5Ew | pufa5Ew | -0.208 | 0.101 | -0.406 | -0.010 | 0.040 | 0.000 | 0.711 |
| BMI | rs543874 | fb_veg_kc | fruit | -0.053 | 0.026 | -0.103 | -0.003 | 0.040 | 0.922 | 0.422 |
| BMI | rs12885454 | mufa5Ew | cho5Ew | -0.091 | 0.044 | -0.177 | -0.004 | 0.040 | 0.000 | 0.586 |
| T2D | rs10278336 | fb_veg_kc | fruit | -0.055 | 0.027 | -0.107 | -0.002 | 0.040 | 19.091 | 0.279 |
| BMI | rs10733682 | cho5Ew | prota5Ew | 0.038 | 0.019 | 0.002 | 0.075 | 0.040 | 15.642 | 0.307 |
| BMI | rs758747 | prota5Ew | pufa5Ew | -0.093 | 0.045 | -0.181 | -0.004 | 0.040 | 0.000 | 0.508 |
| BMI | rs10733682 | cho5Ew | sfa5Ew | 0.036 | 0.018 | 0.002 | 0.071 | 0.041 | 6.244 | 0.382 |
| BMI | rs758747 | prota5Ew | sfa5Ew | -0.093 | 0.045 | -0.182 | -0.004 | 0.041 | 0.000 | 0.506 |
| T2D | rs2943640 | protp5Ew | mufa5Ew | 0.243 | 0.119 | 0.010 | 0.477 | 0.041 | 0.000 | 0.602 |
| BMI | rs1516725 | mufa5Ew | cho5Ew | -0.130 | 0.064 | -0.255 | -0.005 | 0.041 | 0.000 | 0.857 |
| BMI | rs2112347 | protp5Ew | sfa5Ew | -0.206 | 0.101 | -0.404 | -0.008 | 0.041 | 0.000 | 0.755 |
| BMI | rs758747 | prot5Ew | fat5Ew | -0.100 | 0.049 | -0.197 | -0.004 | 0.041 | 0.000 | 0.747 |
| IR | rs308971 | prota5Ew | pufa5Ew | -0.203 | 0.099 | -0.397 | -0.008 | 0.041 | 42.777 | 0.093 |
| IR | rs731839 | protp5Ew | pufa5Ew | -0.249 | 0.122 | -0.489 | -0.010 | 0.041 | 0.000 | 0.757 |
| T2D | rs3802177 | cho5Ew | prota5Ew | 0.047 | 0.023 | 0.002 | 0.092 | 0.041 | 7.864 | 0.369 |
| BMI | rs16851483 | cho5Ew | prota5Ew | -0.073 | 0.036 | -0.143 | -0.003 | 0.042 | 0.000 | 0.669 |
| T2D | rs780094 | mufa5Ew | cho5Ew | -0.112 | 0.055 | -0.219 | -0.004 | 0.042 | 8.556 | 0.364 |
| BMI | rs9540493 | cho5Ew | protp5Ew | -0.038 | 0.019 | -0.074 | -0.001 | 0.042 | 11.755 | 0.339 |
| BMI | rs2365389 | protp5Ew | pufa5Ew | -0.281 | 0.138 | -0.551 | -0.011 | 0.042 | 39.244 | 0.117 |
| BMI | rs13078960 | pufa5Ew | prot5Ew | 0.157 | 0.077 | 0.006 | 0.308 | 0.042 | 0.000 | 0.564 |
| IR | rs7973683 | cho5Ew | fat5Ew | -0.040 | 0.019 | -0.078 | -0.001 | 0.042 | 0.000 | 0.694 |
| BMI | rs9540493 | cho5Ew | sfa5Ew | -0.036 | 0.018 | -0.071 | -0.001 | 0.042 | 5.137 | 0.391 |
| T2D | rs11634397 | pufa5Ew | cho5Ew | -0.161 | 0.080 | -0.317 | -0.005 | 0.043 | 0.000 | 0.849 |
| T2D | rs2943640 | protp5Ew | sfa5Ew | 0.241 | 0.119 | 0.008 | 0.475 | 0.043 | 0.000 | 0.575 |
| IR | rs6066149 | protp5Ew | prota5Ew | -0.274 | 0.135 | -0.539 | -0.009 | 0.043 | 0.000 | 0.532 |
| BMI | rs1516725 | mufa5Ew | sfa5Ew | -0.130 | 0.064 | -0.255 | -0.004 | 0.043 | 0.000 | 0.848 |
| BMI | rs2365389 | protp5Ew | fat5Ew | -0.263 | 0.130 | -0.517 | -0.008 | 0.043 | 33.496 | 0.161 |
| T2D | rs780094 | pufa5Ew | sfa5Ew | 0.204 | 0.101 | 0.006 | 0.402 | 0.043 | 31.397 | 0.177 |
| BMI | rs13107325 | prota5Ew | fat5Ew | -0.162 | 0.080 | -0.319 | -0.005 | 0.043 | 0.000 | 0.694 |
| BMI | rs2365389 | protp5Ew | mufa5Ew | -0.278 | 0.138 | -0.549 | -0.008 | 0.044 | 39.794 | 0.114 |
| IR | rs11130329 | fb_kcalw | fruitw | 0.035 | 0.017 | 0.001 | 0.068 | 0.044 | 38.063 | 0.126 |
| BMI | rs12885454 | mufa5Ew | sfa5Ew | -0.089 | 0.044 | -0.176 | -0.002 | 0.044 | 0.000 | 0.534 |
| IR | rs1045241 | prota5Ew | sfa5Ew | 0.105 | 0.052 | 0.003 | 0.208 | 0.044 | 0.000 | 0.633 |
| BMI | rs13107325 | prota5Ew | mufa5Ew | -0.161 | 0.080 | -0.318 | -0.004 | 0.045 | 0.000 | 0.674 |
| BMI | rs2033732 | protp5Ew | prota5Ew | -0.241 | 0.120 | -0.477 | -0.006 | 0.045 | 0.000 | 0.590 |
| IR | rs3864041 | protp5Ew | sfa5Ew | 0.241 | 0.120 | 0.006 | 0.477 | 0.045 | 0.000 | 0.875 |
| BMI | rs2365389 | protp5Ew | sfa5Ew | -0.280 | 0.140 | -0.553 | -0.006 | 0.045 | 40.412 | 0.109 |
| IR | rs731839 | protp5Ew | fat5Ew | -0.244 | 0.122 | -0.483 | -0.005 | 0.045 | 0.000 | 0.770 |
| BMI | rs13078960 | pufa5Ew | sfa5Ew | 0.155 | 0.077 | 0.003 | 0.307 | 0.045 | 0.000 | 0.727 |
| BMI | rs10733682 | cho5Ew | protp5Ew | 0.036 | 0.018 | 0.001 | 0.071 | 0.046 | 9.665 | 0.355 |
| IR | rs3864041 | protp5Ew | pufa5Ew | 0.240 | 0.120 | 0.004 | 0.475 | 0.046 | 0.000 | 0.793 |
| BMI | rs2820292 | pufa5Ew | cho5Ew | 0.133 | 0.067 | 0.002 | 0.265 | 0.046 | 0.000 | 0.750 |
| T2D | rs849135 | prot5Ew | fat5Ew | 0.101 | 0.051 | 0.002 | 0.201 | 0.046 | 0.000 | 0.670 |
| T2D | rs11634397 | pufa5Ew | sfa5Ew | -0.158 | 0.079 | -0.313 | -0.002 | 0.047 | 0.000 | 0.820 |
| T2D | rs3802177 | protp5Ew | mufa5Ew | 0.250 | 0.126 | 0.003 | 0.498 | 0.048 | 0.000 | 0.974 |
| BMI | rs2820292 | pufa5Ew | sfa5Ew | 0.133 | 0.067 | 0.001 | 0.264 | 0.048 | 0.000 | 0.803 |
| BMI | rs16851483 | cho5Ew | prot5Ew | -0.071 | 0.036 | -0.141 | -0.001 | 0.048 | 0.000 | 0.656 |
| BMI | rs2033732 | protp5Ew | sfa5Ew | -0.238 | 0.121 | -0.475 | -0.002 | 0.048 | 0.000 | 0.474 |
| T2D | rs11257655 | pufa5Ew | sfa5Ew | 0.251 | 0.127 | 0.002 | 0.501 | 0.048 | 37.663 | 0.129 |
| BMI | rs3736485 | cho5Ew | prot5Ew | -0.034 | 0.017 | -0.069 | 0.000 | 0.049 | 5.429 | 0.388 |
| BMI | rs16851483 | cho5Ew | fat5Ew | -0.071 | 0.036 | -0.142 | 0.000 | 0.049 | 0.000 | 0.808 |
| IR | rs7973683 | cho5Ew | prota5Ew | -0.039 | 0.020 | -0.077 | 0.000 | 0.049 | 0.000 | 0.663 |
| BMI | rs977747 | prota5Ew | sfa5Ew | -0.096 | 0.049 | -0.191 | 0.000 | 0.049 | 22.840 | 0.248 |
| IR | rs1045241 | prota5Ew | mufa5Ew | 0.103 | 0.052 | 0.000 | 0.206 | 0.049 | 0.000 | 0.687 |
| BMI | rs11727676 | protp5Ew | fat5Ew | 0.379 | 0.193 | 0.001 | 0.757 | 0.049 | 0.000 | 0.713 |
| BMI | rs758747 | prot5Ew | sfa5Ew | -0.097 | 0.050 | -0.194 | 0.000 | 0.050 | 0.000 | 0.708 |
| T2D | rs11257655 | pufa5Ew | cho5Ew | 0.256 | 0.130 | 0.000 | 0.511 | 0.050 | 39.751 | 0.114 |

## *Appendix H: Individual SNP and food and beverage interactions and incident T2D, with p-value for interaction<0.05*

Please note that these interactions are based on model 4 detailed in Table 6-4.

GRS: genetic risk score
snp: single nucleotide polymorphism
beta: beta coefficient
se: standard error
lci: lower confidence interval
uci: upper confidence interval
p_int: pvalue for interaction
i_sq: I squared for heterogeneity
p_het: pvalue for heterogeneity
T2D: type 2 diabetes
IR: insulin resistance
BMI: body mass index
ssb: sugar sweetened beverage

Threshold for p value for interaction after Bonferroni correction< p<1.71E-05
Please note:
All foods and beverages are per 100g or 100ml, except for nuts and seeds which is per 30g
This analysis includes 7 of the 8 countries. France was excluded because of the spurious interactions arising, likely due to low sample size.
We also checked the analysis by merging France with Germany and the results were very similar.

| GRS | snp | Food or beverage | beta | se | lci | uci | p_int | i_sq | p_het |
|---|---|---|---|---|---|---|---|---|---|
| T2D | rs6878122 | wholegrain | -0.201 | 0.052 | -0.303 | -0.098 | 1.269E-04 | 0.000 | 0.499 |
| BMI | rs2820292 | coffee | 0.027 | 0.007 | 0.012 | 0.042 | 3.028E-04 | 0.000 | 0.449 |
| T2D | rs11063069 | tea | 0.045 | 0.013 | 0.020 | 0.071 | 4.941E-04 | 0.000 | 0.704 |
| T2D | rs10278336 | green leafy vegetables | -0.261 | 0.083 | -0.423 | -0.098 | 0.002 | 0.000 | 0.817 |
| BMI | rs10132280 | fruit | -0.047 | 0.015 | -0.077 | -0.018 | 0.002 | 0.000 | 0.696 |
| BMI | rs13107325 | nuts and seeds | 0.840 | 0.271 | 0.310 | 1.371 | 0.002 | 0.000 | 0.440 |
| IR | rs1045241 | fruit | -0.057 | 0.018 | -0.093 | -0.021 | 0.002 | 0.000 | 0.805 |
| BMI | rs3736485 | nuts and seeds | -0.375 | 0.124 | -0.619 | -0.131 | 0.003 | 0.000 | 0.766 |
| T2D | rs459193 | tea | -0.038 | 0.013 | -0.063 | -0.013 | 0.003 | 0.000 | 0.880 |
| T2D | rs7177055 | ssb | -0.071 | 0.024 | -0.119 | -0.024 | 0.003 | 0.000 | 0.517 |
| BMI | rs2033732 | rice | -0.393 | 0.134 | -0.655 | -0.130 | 0.003 | 0.000 | 0.956 |
| T2D | rs2334499 | ssb | 0.088 | 0.031 | 0.027 | 0.149 | 0.005 | 34.113 | 0.168 |
| BMI | rs1441264 | ssb | 0.061 | 0.022 | 0.018 | 0.103 | 0.005 | 0.000 | 0.850 |
| IR | rs11577194 | rice | -0.352 | 0.127 | -0.600 | -0.103 | 0.005 | 0.000 | 0.536 |
| T2D | rs6795735 | fish | 0.567 | 0.207 | 0.162 | 0.972 | 0.006 | 0.000 | 0.605 |
| BMI | rs3849570 | egg and egg products | 0.402 | 0.147 | 0.113 | 0.691 | 0.006 | 0.000 | 0.704 |
| BMI | rs4740619 | fish | 0.480 | 0.178 | 0.131 | 0.829 | 0.007 | 0.000 | 0.815 |
| T2D | rs7177055 | tea | 0.033 | 0.012 | 0.009 | 0.057 | 0.007 | 0.250 | 0.422 |
| BMI | rs7164727 | wholegrain | 0.121 | 0.046 | 0.031 | 0.212 | 0.009 | 1.809 | 0.411 |
| T2D | rs7756992 | root vegetables | -0.316 | 0.121 | -0.553 | -0.079 | 0.009 | 1.767 | 0.411 |
| BMI | rs3849570 | fish | 0.477 | 0.185 | 0.114 | 0.841 | 0.010 | 0.000 | 0.987 |
| IR | rs8101064 | processed meat | -0.612 | 0.241 | -1.084 | -0.140 | 0.011 | 0.000 | 0.684 |
| IR | rs754814 | green leafy vegetables | -0.223 | 0.088 | -0.396 | -0.050 | 0.011 | 0.000 | 0.924 |
| BMI | rs205262 | coffee | 0.020 | 0.008 | 0.004 | 0.037 | 0.012 | 0.000 | 0.470 |
| T2D | rs7177055 | red meat | -0.222 | 0.089 | -0.397 | -0.047 | 0.013 | 0.000 | 0.942 |
| BMI | rs2836754 | fruit | 0.037 | 0.015 | 0.008 | 0.066 | 0.013 | 0.000 | 0.796 |
| BMI | rs7599312 | ssb | 0.053 | 0.022 | 0.011 | 0.095 | 0.014 | 0.000 | 0.596 |
| BMI | rs16907751 | green leafy vegetables | -0.282 | 0.116 | -0.510 | -0.055 | 0.015 | 0.000 | 0.936 |
| IR | rs4804311 | legumes | -0.451 | 0.186 | -0.816 | -0.085 | 0.016 | 0.000 | 0.479 |
| T2D | rs10830963 | nuts and seeds | -0.371 | 0.154 | -0.673 | -0.070 | 0.016 | 0.000 | 0.898 |
| T2D | rs13389219 | ssb | 0.065 | 0.027 | 0.012 | 0.119 | 0.016 | 23.648 | 0.249 |
| T2D | rs849135 | fermented dairy | -0.075 | 0.031 | -0.137 | -0.014 | 0.017 | 0.000 | 0.457 |
| BMI | rs9914578 | nuts and seeds | -0.385 | 0.163 | -0.705 | -0.066 | 0.018 | 0.000 | 0.697 |
| BMI | rs7599312 | tea | 0.024 | 0.010 | 0.004 | 0.044 | 0.019 | 0.000 | 0.675 |
| BMI | rs9374842 | ssb | 0.063 | 0.027 | 0.011 | 0.116 | 0.019 | 22.663 | 0.256 |
| T2D | rs11717195 | egg and egg products | 0.456 | 0.195 | 0.075 | 0.838 | 0.019 | 2.037 | 0.409 |
| BMI | rs2080454 | red meat | -0.168 | 0.072 | -0.310 | -0.026 | 0.020 | 0.000 | 0.763 |
| T2D | rs1496653 | ssb | -0.077 | 0.033 | -0.142 | -0.012 | 0.021 | 17.238 | 0.298 |
| BMI | rs3810291 | legumes | 0.363 | 0.157 | 0.055 | 0.670 | 0.021 | 7.275 | 0.370 |
| IR | rs10195252 | ssb | 0.057 | 0.025 | 0.008 | 0.106 | 0.022 | 13.853 | 0.324 |
| IR | rs6066149 | tea | 0.052 | 0.023 | 0.007 | 0.096 | 0.024 | 43.827 | 0.099 |
| BMI | rs13078960 | nuts and seeds | 0.324 | 0.144 | 0.042 | 0.607 | 0.024 | 0.000 | 0.621 |
| BMI | rs9540493 | egg and egg products | 0.321 | 0.142 | 0.041 | 0.600 | 0.024 | 0.000 | 0.656 |
| T2D | rs11651052 | ssb | -0.057 | 0.025 | -0.106 | -0.007 | 0.025 | 5.262 | 0.387 |
| IR | rs645040 | fish | -0.618 | 0.275 | -1.158 | -0.079 | 0.025 | 14.291 | 0.321 |
| BMI | rs6091540 | root vegetables | 0.235 | 0.104 | 0.030 | 0.439 | 0.025 | 0.000 | 0.997 |
| T2D | rs243088 | legumes | -0.314 | 0.140 | -0.588 | -0.039 | 0.025 | 0.000 | 0.469 |
| BMI | rs12566985 | nuts and seeds | 0.286 | 0.128 | 0.035 | 0.538 | 0.025 | 0.000 | 0.633 |
| BMI | rs16907751 | processed meat | -0.310 | 0.139 | -0.582 | -0.038 | 0.025 | 15.906 | 0.309 |
| T2D | rs10401969 | legumes | 0.675 | 0.304 | 0.078 | 1.271 | 0.027 | 0.000 | 0.734 |
| IR | rs683135 | fermented dairy | 0.094 | 0.043 | 0.011 | 0.178 | 0.027 | 31.238 | 0.190 |
| BMI | rs12286929 | egg and egg products | 0.307 | 0.139 | 0.035 | 0.580 | 0.027 | 0.000 | 0.600 |
| IR | rs132985 | fermented dairy | 0.066 | 0.030 | 0.007 | 0.124 | 0.028 | 0.000 | 0.843 |
| IR | rs683135 | tea | 0.037 | 0.017 | 0.004 | 0.070 | 0.028 | 23.774 | 0.248 |
| BMI | rs2820292 | legumes | 0.550 | 0.250 | 0.060 | 1.040 | 0.028 | 37.236 | 0.158 |
| T2D | rs11063069 | rice | 0.342 | 0.156 | 0.036 | 0.648 | 0.028 | 0.000 | 0.672 |
| T2D | rs6878122 | green leafy vegetables | -0.176 | 0.081 | -0.335 | -0.018 | 0.029 | 0.000 | 0.835 |
| BMI | rs9641123 | ssb | 0.048 | 0.022 | 0.005 | 0.091 | 0.029 | 0.000 | 0.466 |

| GRS | snp | Food or beverage | beta | se | lci | uci | p_int | i_sq | p_het |
|---|---|---|---|---|---|---|---|---|---|
| T2D | rs13233731 | green leafy vegetables | 0.284 | 0.132 | 0.025 | 0.543 | 0.031 | 12.941 | 0.331 |
| T2D | rs7177055 | coffee | -0.020 | 0.009 | -0.037 | -0.002 | 0.032 | 0.000 | 0.450 |
| BMI | rs2075650 | processed meat | -0.246 | 0.115 | -0.472 | -0.021 | 0.032 | 0.000 | 0.507 |
| T2D | rs1552224 | wholegrain | 0.149 | 0.070 | 0.013 | 0.285 | 0.032 | 0.000 | 0.905 |
| BMI | rs543874 | ssb | 0.056 | 0.026 | 0.005 | 0.107 | 0.033 | 0.000 | 0.866 |
| BMI | rs3849570 | red meat | 0.154 | 0.072 | 0.013 | 0.296 | 0.033 | 0.000 | 0.467 |
| BMI | rs9914578 | green leafy vegetables | -0.268 | 0.126 | -0.514 | -0.022 | 0.033 | 6.431 | 0.379 |
| BMI | rs6804842 | egg and egg products | 0.301 | 0.141 | 0.024 | 0.577 | 0.033 | 0.000 | 0.802 |
| T2D | rs2261181 | nuts and seeds | 0.911 | 0.428 | 0.072 | 1.750 | 0.033 | 60.030 | 0.020 |
| BMI | rs3101336 | egg and egg products | -0.307 | 0.144 | -0.590 | -0.025 | 0.033 | 0.000 | 0.808 |
| T2D | rs1801282 | fruit | 0.065 | 0.030 | 0.005 | 0.124 | 0.034 | 15.950 | 0.308 |
| BMI | rs1928295 | ssb | -0.042 | 0.020 | -0.080 | -0.003 | 0.034 | 0.000 | 0.747 |
| BMI | rs10132280 | root vegetables | 0.230 | 0.109 | 0.017 | 0.444 | 0.034 | 0.000 | 0.732 |
| IR | rs9881942 | egg and egg products | -0.340 | 0.161 | -0.655 | -0.024 | 0.035 | 0.000 | 0.756 |
| IR | rs731839 | fermented dairy | 0.069 | 0.033 | 0.005 | 0.133 | 0.035 | 0.000 | 0.804 |
| BMI | rs4256980 | tea | 0.031 | 0.015 | 0.002 | 0.060 | 0.035 | 25.466 | 0.234 |
| T2D | rs1359790 | ssb | 0.122 | 0.058 | 0.008 | 0.236 | 0.035 | 76.633 | 0.000 |
| BMI | rs2365389 | wholegrain | -0.090 | 0.043 | -0.174 | -0.006 | 0.036 | 0.000 | 0.620 |
| T2D | rs6795735 | red meat | 0.169 | 0.080 | 0.011 | 0.326 | 0.036 | 0.000 | 0.785 |
| BMI | rs12401738 | ssb | 0.047 | 0.023 | 0.003 | 0.091 | 0.036 | 8.460 | 0.364 |
| IR | rs498313 | green leafy vegetables | -0.237 | 0.113 | -0.459 | -0.015 | 0.037 | 4.368 | 0.393 |
| BMI | rs1167827 | processed meat | 0.224 | 0.108 | 0.013 | 0.435 | 0.037 | 38.178 | 0.138 |
| IR | rs498313 | fish | -0.455 | 0.220 | -0.886 | -0.025 | 0.038 | 0.000 | 0.779 |
| T2D | rs10830963 | root vegetables | -0.230 | 0.112 | -0.449 | -0.012 | 0.039 | 0.000 | 0.796 |
| BMI | rs1558902 | legumes | -0.262 | 0.127 | -0.511 | -0.013 | 0.039 | 0.000 | 0.640 |
| BMI | rs17094222 | fish | 0.549 | 0.266 | 0.027 | 1.071 | 0.039 | 14.466 | 0.319 |
| T2D | rs10278336 | rice | 0.265 | 0.129 | 0.013 | 0.517 | 0.039 | 0.000 | 0.571 |
| BMI | rs1516725 | tea | -0.029 | 0.014 | -0.057 | -0.001 | 0.040 | 0.000 | 0.872 |
| T2D | rs12427353 | wholegrain | -0.128 | 0.062 | -0.250 | -0.006 | 0.040 | 0.000 | 0.553 |
| IR | rs2699429 | fish | 0.416 | 0.203 | 0.019 | 0.814 | 0.040 | 0.000 | 0.482 |
| IR | rs2126259 | legumes | -0.509 | 0.248 | -0.995 | -0.024 | 0.040 | 0.000 | 0.691 |
| BMI | rs7243357 | ssb | -0.056 | 0.027 | -0.109 | -0.002 | 0.040 | 0.000 | 0.620 |
| BMI | rs4256980 | red meat | -0.148 | 0.072 | -0.289 | -0.006 | 0.040 | 0.000 | 0.784 |
| IR | rs17402950 | wholegrain | -0.252 | 0.124 | -0.495 | -0.010 | 0.041 | 12.991 | 0.331 |
| IR | rs3861397 | fermented dairy | 0.065 | 0.032 | 0.002 | 0.128 | 0.042 | 0.000 | 0.707 |
| BMI | rs11165643 | green leafy vegetables | -0.155 | 0.076 | -0.304 | -0.005 | 0.043 | 0.000 | 0.669 |
| BMI | rs2176040 | nuts and seeds | 0.270 | 0.133 | 0.009 | 0.531 | 0.043 | 0.000 | 0.453 |
| IR | rs132985 | egg and egg products | 0.422 | 0.208 | 0.014 | 0.831 | 0.043 | 29.194 | 0.205 |
| BMI | rs11126666 | red meat | -0.161 | 0.079 | -0.316 | -0.005 | 0.043 | 0.000 | 0.774 |
| BMI | rs17001654 | fish | -0.493 | 0.244 | -0.972 | -0.015 | 0.043 | 0.000 | 0.928 |
| IR | rs498313 | nuts and seeds | 0.462 | 0.230 | 0.011 | 0.914 | 0.045 | 42.676 | 0.106 |
| IR | rs2126259 | fish | 0.698 | 0.348 | 0.016 | 1.380 | 0.045 | 0.000 | 0.442 |
| IR | rs11130329 | legumes | 0.971 | 0.483 | 0.024 | 1.919 | 0.045 | 50.035 | 0.075 |
| BMI | rs12429545 | coffee | -0.020 | 0.010 | -0.040 | 0.000 | 0.045 | 0.000 | 0.424 |
| IR | rs6822892 | fish | -0.420 | 0.210 | -0.830 | -0.009 | 0.045 | 0.000 | 0.847 |
| T2D | rs17168486 | tea | 0.057 | 0.029 | 0.001 | 0.113 | 0.045 | 51.205 | 0.056 |
| BMI | rs6091540 | processed meat | -0.170 | 0.085 | -0.336 | -0.003 | 0.046 | 0.000 | 0.891 |
| T2D | rs10923931 | fermented dairy | 0.137 | 0.069 | 0.002 | 0.271 | 0.046 | 37.788 | 0.140 |
| BMI | rs11191560 | ssb | 0.073 | 0.037 | 0.001 | 0.146 | 0.047 | 8.405 | 0.364 |
| T2D | rs7202877 | rice | -0.546 | 0.276 | -1.086 | -0.006 | 0.048 | 33.410 | 0.173 |
| IR | rs132985 | fish | -0.515 | 0.261 | -1.027 | -0.004 | 0.048 | 23.666 | 0.249 |
| IR | rs2434612 | red meat | -0.191 | 0.097 | -0.380 | -0.001 | 0.048 | 0.000 | 0.647 |
| BMI | rs12940622 | green leafy vegetables | -0.153 | 0.078 | -0.305 | -0.001 | 0.049 | 0.000 | 0.844 |
| BMI | rs12446632 | coffee | -0.021 | 0.011 | -0.042 | 0.000 | 0.049 | 0.000 | 0.576 |
| BMI | rs3101336 | rice | 0.231 | 0.117 | 0.001 | 0.461 | 0.049 | 0.000 | 0.526 |
| T2D | rs10401969 | tea | 0.035 | 0.018 | 0.000 | 0.069 | 0.049 | 0.965 | 0.417 |
| T2D | rs13233731 | red meat | 0.215 | 0.110 | 0.000 | 0.431 | 0.050 | 40.426 | 0.122 |
| BMI | rs7899106 | ssb | 0.147 | 0.075 | 0.000 | 0.294 | 0.050 | 52.242 | 0.051 |
| IR | rs6937438 | fermented dairy | 0.068 | 0.035 | 0.000 | 0.135 | 0.050 | 0.000 | 0.551 |
| IR | rs7005992 | ssb | 0.085 | 0.044 | 0.000 | 0.171 | 0.050 | 39.576 | 0.128 |
| IR | rs2126259 | green leafy vegetables | -1.086 | 0.555 | -2.174 | 0.001 | 0.050 | 69.477 | 0.003 |
| BMI | rs2207139 | wholegrain | -0.140 | 0.072 | -0.281 | 0.000 | 0.050 | 31.884 | 0.185 |
| BMI | rs2836754 | fish | -0.379 | 0.194 | -0.759 | 0.001 | 0.050 | 0.000 | 0.453 |

## *Appendix I: Top 5 SNPs for each macronutrient from a genome-environment-wide- interaction-study for the risk of developing Type 2 Diabetes*

**Table: Top five hits from the GEWIS of each respective macronutrient intake and incident T2D (EPIC-InterAct)**

| Macro-nutrient | SNP | Chromosome: position: effect allele: other allele | Gene (in/near) | Effect allele freq-uency | β | SE | p |
|---|---|---|---|---|---|---|---|
| Carbohydrate | rs3112356 | 7:135396031_A_G | *SLC13A4* | 0.23 | 0.12 | 0.03 | 2.42E-06 |
| | rs8138283 | 22:49841479_A_G | *C22orf34* | 0.66 | 0.10 | 0.02 | 5.14E-06 |
| | rs116950933 | 8:145959416_G_A | *LOC107986986* | 0.02 | -0.55 | 0.12 | 5.89E-06 |
| | rs2359460 | 3:23111674_A_G | intergenic | 0.91 | 0.15 | 0.03 | 6.17E-06 |
| | rs4902166 | 14:63123208_G_T | intergenic | 0.85 | -0.14 | 0.03 | 6.87E-06 |
| Protein | rs61810167 | 1:163082778_T_C | *RGS5* | 0.99 | -1.97 | 0.39 | 3.93E-07 |
| | rs186453404 | 14:87109364_T_C | intergenic | 0.99 | -0.97 | 0.20 | 1.44E-06 |
| | rs10999966 | 10:73468137_G_A | *CDH23* | 0.04 | 0.50 | 0.11 | 1.67E-06 |
| | rs141413593 | 17:75763769_C_G | intergenic | 0.96 | -0.54 | 0.11 | 2.35E-06 |
| | rs11786114 | 8:102321001_C_T | *RNU7-67P* | 0.03 | -0.63 | 0.13 | 2.84E-06 |
| SFA | rs73652097 | 9:81385044_T_C | intergenic | 0.88 | -0.32 | 0.07 | 1.41E-06 |
| | rs7725760 | 5:39556805_A_G | intergenic | 0.78 | -0.22 | 0.05 | 1.97E-06 |
| | rs12338311 | 9:81385696_A_G | intergenic | 0.88 | -0.31 | 0.07 | 1.98E-06 |
| | rs142865342 | 21:46284404_C_T | *PTTG1IP* | 0.01 | 0.90 | 0.19 | 2.30E-06 |
| | rs2962495 | 5:39574354_A_G | *CCDC11P1* | 0.23 | 0.22 | 0.05 | 2.96E-06 |
| PUFA | rs11003372 | 10:54840408_G_A | intergenic | 0.22 | 0.62 | 0.13 | 1.33E-06 |
| | rs12099307 | 11:125990750_G_T | intergenic | 0.07 | 0.68 | 0.15 | 4.99E-06 |
| | rs12099306 | 11:125990749_G_C | intergenic | 0.07 | 0.68 | 0.15 | 5.06E-06 |
| | rs146302320 | 8:140889848_G_A | *TRAPPC9* | 0.01 | 1.02 | 0.23 | 6.27E-06 |
| | rs72793267 | 10:52965186_C_T | *PRKG1* | 0.05 | 1.02 | 0.23 | 6.77E-06 |
| Cereal fibre | rs4689848 | 4:7751894_C_T | *AFAP1-AS1* | 0.33 | -0.07 | 0.01 | 1.29E-07 |
| | rs4689849 | 4:7752174_C_T | *AFAP1-AS1* | 0.34 | -0.07 | 0.01 | 2.32E-07 |
| | rs11721409 | 4:7753417_C_T | *AFAP1-AS1* | 0.34 | -0.07 | 0.01 | 7.30E-07 |
| | rs4018903 | 3:193900347_A_G | intergenic | 0.54 | -0.08 | 0.02 | 7.64E-07 |
| | rs6855176 | 4:7753403_C_G | *AFAP1-AS1* | 0.34 | -0.07 | 0.01 | 7.74E-07 |

Effect estimates for the interaction between macronutrient intake (per 5% total energy intake or per g/1000kcal for cereal fibre) and per risk allele of the respective SNP on incident T2D. Abbreviations: SFA- saturated fat, MUFA- monounsaturated fat, PUFA- polyunsaturated fat, N- number, SE- standard error, p- value for interaction